

PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF USING THE SAME

RELATED APPLICATIONS

5 This application claims priority to U.S.S.N. 60/256,025, filed December 15, 2000;
U.S.S.N. 60/265,163, filed January 30, 2001; U.S.S.N. 60/272,929, filed March 2, 2001;
U.S.S.N. 60/274,864, filed March 9, 2001; U.S.S.N. 60/276,688, filed March 16, 2001;
U.S.S.N. 60/277,880, filed March 22, 2001; U.S.S.N. 60/286,409, filed April 25, 2001;
U.S.S.N. 60/309,246, filed July 31, 2001, and U.S.S.N. 60/315,600, filed August 29, 2001
10 each of which is incorporated by reference in its entirety.

FIELD OF THE INVENTION

The invention relates to polynucleotides and the polypeptides encoded by such
polynucleotides, as well as vectors, host cells, antibodies and recombinant methods for
producing the polypeptides and polynucleotides, as well as methods for using the same.

15 BACKGROUND OF THE INVENTION

The present invention is based in part on nucleic acids encoding proteins that are new
members of the following protein families: Potassium Channel-like, Galanin Receptor Type 1
(GAL1-R) (GALR1)-like, P2Y Purinoceptor-1-like, LOMP-like, Epidermal Growth Factor-
like, Hyaluronan Mediated Motility Receptor-like, Serpin-like, B7 family-like and Acyl CoA
20 Dehydrogenase-like. More particularly, the invention relates to nucleic acids encoding novel
polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for
producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding
25 novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as NOVX,
or NOV1, NOV2, NOV3, NOV4, NOV5, NOV6, NOV7, NOV8 and NOV9 nucleic acids and
polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs
and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or

polypeptide sequences.

In one aspect, the invention provides an isolated NOVX nucleic acid molecule encoding a NOVX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41. In some embodiments, the NOVX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41) or a complement of said oligonucleotide. Also included in the invention are substantially purified NOVX polypeptides (SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42). In certain embodiments, the NOVX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

Also included in the invention is a method of detecting the presence of a NOVX nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, behavioral disorders, addiction, anxiety, pain, actinic keratosis, acne, hair growth diseases, alopecia, pigmentation disorders, endocrine disorders, connective tissue disorders, such as severe neonatal Marfan syndrome, dominant ectopia lentis, familial ascending aortic aneurysm, isolated skeletal features of Marfan syndrome, Shprintzen-Goldberg syndrome, genodermatoses, contractural arachnodactyly, inflammatory disorders such as osteo- and rheumatoid-arthritis, inflammatory bowel disease, Crohn's disease; immunological disorders, AIDS; cancers including but not limited to lung cancer, colon cancer, Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer, leukemia or pancreatic cancer; blood disorders; asthma; psoriasis; vascular disorders, hypertension, skin disorders, renal disorders including Alport syndrome, immunological disorders, tissue injury, fibrosis disorders, bone diseases, Ehlers-Danlos syndrome type VI, VII, type IV, S-linked cutis laxa and Ehlers-Danlos syndrome type V, osteogenesis imperfecta,

Neurologic diseases, Brain and/or autoimmune disorders like encephalomyelitis, neurodegenerative disorders, immune disorders, hematopoietic disorders, muscle disorders, inflammation and wound repair, bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, benign prostatic hypertrophy, arthrogryposis multiplex congenita, osteogenesis imperfecta, keratoconus, scoliosis, duodenal atresia, esophageal atresia, intestinal malrotation, Pancreatitis, Obesity Systemic lupus erythematosus, Autoimmune disease, Emphysema, Scleroderma, allergy, ARDS, Neuroprotection, Fertility Myasthenia gravis, Diabetes, obesity, Growth and reproductive disorders Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host, Adrenoleukodystrophy, Congenital Adrenal Hyperplasia, Endometriosis, Xerostomia, Ulcers, Cirrhosis, Transplantation, Diverticular disease, Hirschsprung's disease, Appendicitis, Arthritis, Ankylosing spondylitis, Tendinitis, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, erythematosus, Renal tubular acidosis, IgA nephropathy, anorexia, bulimia, psychotic disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease and/or other pathologies and disorders of the like.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in gene therapy, and NOVX may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

The invention further includes a method for screening for a modulator of disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The method includes contacting a test compound with a NOVX polypeptide and determining if the test compound binds to said NOVX polypeptide. Binding

of the test compound to the NOVX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a NOVX nucleic acid. Expression or activity of NOVX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal which recombinantly-expresses NOVX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of NOVX polypeptide in both the test animal and the control animal is compared. A change in the activity of NOVX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide, a NOVX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the amount of the NOVX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the NOVX polypeptide present in a control sample. An alteration in the level of the NOVX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a NOVX polypeptide, a NOVX nucleic acid, or a NOVX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identity the cellular receptors and downstream effectors of the invention by any one of a number of techniques

commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

NOVX nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOVX substances for use in therapeutic or
5 diagnostic methods. These NOVX antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOVX proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These NOVX proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding
10 of pathology of the disease and development of new drug targets for various disorders.

The NOVX nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic,
15 diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration in vivo and in vitro of all tissues and cell types composing (but not limited to) those defined here.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention
20 belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods,
25 and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby.
30 Included in the invention are the novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to

autoimmune disease, asthma, emphysema, scleroderma, allergy and/or ARDS and/or other pathologies/disorders.

NOV2 is homologous to a Galanin Receptor Type 1 (GAL1-R) (GALR1)-like family of proteins. Thus NOV2 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain and/or neuroprotection and/or other pathologies/disorders.

NOV3 is homologous to a family of P2Y Purinoceptor-1-like proteins. Thus, the NOV3 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example: hyperparathyroidism, fertility, endometriosis, Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy and/or ARDS and/or other pathologies/disorders.

NOV4 is homologous to the LOMP-like family of proteins. Thus, NOV4 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated, for example; cancer, developmental and/or neurological disorders. Since experimental evidence using the genetics of *Drosophila*, *C. elegans*, and mice indicates that PDZ proteins are involved in the regulation of epithelial cell growth, differentiation, and morphogenetic movements during development, as well as in the interactions among the components of synaptic junctions and/or other pathologies/disorders.

NOV5 is homologous to the Epidermal Growth Factor-like protein family. Thus NOV5 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example: Agammaglobulinemia, type 2, X-linked; Aicardi syndrome; Craniofrontonasal dysplasia; Deafness, X-linked 6, sensorineural; Goiter, multinodular, 2; Mental retardation, X-linked nonspecific, 58; Opitz G syndrome, type I; Partington syndrome II; Simpson-Golabi-Behmel syndrome, type 2; Simpson-Golabi-Behmel syndrome, type 2; Oncogenesis; fertility;

regulation of cell cycle, proliferation and developmental processes and/or other pathologies/disorders.

NOV6 is homologous to the Hyaluronan Mediated Motility Receptor-like family of proteins. Thus NOV6 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example: oncogene-and growth factor-mediated cell locomotion, disorders involving cell locomotion, e.g. tumour invasion, birth defects, acute and chronic inflammatory disorders, Alzheimer's and other forms of dementia, including Parkinson's and Huntington's diseases, AIDS, diabetes, autoimmune diseases, corneal dysplasia and hypertrophies, burns, surgical incisions and adhesions, strokes, breast cancer, Bronchial asthma; Eosinophilia, familial; Muscular dystrophy, limb-girdle, type 2F and multiple sclerosis. They can also be used in e.g. CNS and spinal cord regeneration, contraception and in vitro fertilization and embryo development and/or other pathologies/disorders.

NOV7 is homologous to members of the Serpin-like family of proteins. Thus, the NOV7 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; liver toxicity, cancer, metabolic diseases, inflammation, CNS disorders and/or other pathologies/disorders.

NOV8 is homologous to the B7 family-like family of proteins. Thus, NOV8 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; brain disorders including epilepsy, eating disorders, schizophrenia, ADD, cancer; heart disease, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders, psoriasis, colon cancer, leukemia, AIDS, thalamus disorders, metabolic disorders including diabetes and obesity, lung diseases such as asthma, emphysema, cystic fibrosis, cancer, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer and/or other pathologies/disorders.

NOV9 is homologous to members of the Acyl CoA Dehydrogenase-like family of proteins. Thus, the NOV9 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; obesity, diabetes, cachexia, cancer, inflammation, CNS disorders and SCAD disorders and/or other pathologies/disorders.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, e.g., neurogenesis, cell differentiation, cell proliferation, hematopoiesis, wound healing and angiogenesis.

Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOV1

A disclosed NOV1 nucleic acid of 1953 nucleotides (also referred to as CG50249-01) encoding a novel potassium channel-like protein is shown in Table 1A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 16-18 and ending with a TAA codon at nucleotides 1930-1932. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 1A, and the start and stop codons are in bold letters.

Table 1A. NOV1 Nucleotide Sequence (SEQ ID NO:1).

GTCTGAGTCACAGAGATGGGCAAGATCGAGAACAACGAGAGGGTGATCCTCAATGTGGGGGACCCCGGCACGAAACCTAC
CGCAGCACCTCAAGACCTGCCTGGAACACGCTGGCCCTTCTTGCCTCCTCCGAGCCCCAGGCGACTGCTTGACACG
GCGGGCGACAAGCTGCAGCCGTGCGCGCTCCACTGTGCGCGCGCGGAGAGCGCCCCGCTGTCCCGGGCCAGGCGGC
TGCTTCGAGGGCGGCGCGGCAACTGCAGTTCCGCGGCGGCGAGGCCAGCGACCATCCCGGTGGCGCGCGGAGTTCTTC
TTCCGACCGGCACCCGGGCGTCTTCGCCTATGTGCTCAATTACTACCGCACCGGCAAGCTGCAGTCCCGCGCAGAGCTGTGC
GGGCGCTCTTCGAGGAGGAGCTGGCCTTCTGGGGCATCGACGAGACCGACGTGGAGCCCTGCTGCTGGATGACCTACCGG
CAGCACCGCGCAGCGAGGAGGCGCTGGACATCTTCGAGACCCCGACCTCATTTGGCGGCGACCCCGGCGACGAGGAC
CTGGCGGCAAGAGGCTGGGCATCGAGGACGCGGCGGGCTCGGGGGCCCGCAGCGCAATCTGGCCGCTGGAGGAGGCTG
CAGCCCGCATGTGGCCCTCTTCGAAGACCCCTACTCGTCCAGAGCCCGCAGGTTTATTGCTTTTGCTTCTTATTCTTC
ATCCTGGTTTCAATTACAACITTTTGCCTGGAAACACATGAAGCTTCAATATTGTTAAAAACAAGACAGAACAGTCATC
AATGGCACAAGTGTGTCTACAGTATGAAATTGAACCGGATCCTGCCTTGAGGTATGTAGAAGGAGTGTGTGGTGTGG
TTTACTTTTGAATTTTGTAGTCCGTATTGTTTTCACCCAATAAACTTGAATTCATCAAAAATCTCTTGAATATCATTGAC
TTTGTGGCCATCTACCTTTCTACTTAGAGGTGGGACTCAGTGGGCTGTATCCAAAGCTGCTAAAGATGTGCTTGGCTTC
CTCAGGGTGGTAAGGTTTGTGAGGATCCTGAGAATTTCAAGCTCACCGCCATTTTGTAGGTCTGAGGGTGCTTGGACAT
ACTCTTCGAGCTAGTACTAATGAATTTTGTGCTGATAATTTCTTGGCTCTAGGAGTTTGTATTTGCTACCATGATC
TACTATGCCGAGAGAGTGGGAGCTCAACCTAACGACCCTTCAGCTAGTGAGCACACAGTTCAAAAACATTCCATTGGG
TTCTGGTGGGCTGTAGTGACCATGACTACCTGGGTATGGGGATATGTACCCCAACATGGTCAGGCATGCTGGTGGGA
GCCCTGTGTGCTCTGGCTGGAGTGCTGACAATAGCCATGCCAGTGCCCTGTCAATTGTCAATAATTGGAATGTACTACTCC
TTGGCAATGGCAAAGCAGAACTTCCAAGGAAAAGAAAGACATCCCTCCTGCTCCTCAGGCAAGCTCACCTACTTTT
TGCAAGACAGAATTAATATGGCCTGCAATAGTACACAGAGTGACACATGTCTGGGCAAGACAATCGACTCTGGAACAT
AACAGATCAGTGTATCAGGTGACGACAGTACAGGAAGTGAGCGGCCACTATCACCCCGAGAAAGGCTCCCCATCAGACGC
TCTAGTACCAGAGACAAAACAGAGGGGAAACATGTTTCTACTGACGACAGGTGATTACACGTGTGCTTCTGATGGA
GGGATCAGGAAAGGTTATGAAAAATCCCGAAGCTTAAACAACATAGCGGGCTTGGCAGGCAATGCTCTGAGGCTCTCTCCA
GTAACATCACCTTCAACTCTCTTGTCTCTGAGGCGCTCTCGATCTCCATCCCATCTATCTTGTAAACCAACAACCA
AACTGCATC

The NOV1 nucleic acid sequence maps to chromosome 12 and has 1758 of 1952 bases (90%) identical to a *Rattus norvegicus* K⁺ channel protein (KSHIIIA3) mRNA (gb:GENBANK-ID:RATSHIIC|acc:M84203.1) (E = 0.0). Similarity information was assessed using public nucleotide databases including all GenBank databases and the GeneSeq patent database. Chromosome information was assigned using OMIM and the electronic

northern tool from Curatools to derive the the chromosomal mapping of the SeqCalling assemblies, Genomic clones, and/or EST sequences that were included in the invention.

In all BLAST alignments herein, the "E-value" or "Expect" value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject ("Sbjct") retrieved from the NOV1 BLAST analysis, e.g., *Rattus norvegicus* K⁺ channel protein (KSHIIIA3) mRNA, matched the Query NOV1 sequence purely by chance is 0.0. The Expect value (E) is a parameter that describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.

The Expect value is used as a convenient way to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, e.g., <http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/>. Occasionally, a string of X's or N's will result from a BLAST search. This is a result of automatic filtering of the query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (e.g., "NNNNNNNN") or the letter "X" in protein sequences (e.g., "XXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment. Wootton and Federhen, Methods Enzymol 266:554-571, 1996.

The disclosed NOV1 polypeptide (SEQ ID NO:2) encoded by SEQ ID NO:1 has 638 amino acid residues and is presented in Table 1B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1 does not contain a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000.

Table 1B. Encoded NOV1 protein sequence (SEQ ID NO:2).

```
MGKIENNERVILNVGGTRHETYSRLKTLPGTRLALLASSEPPGDCLTAGDKLQSPFPPLSPPPRAPPLSPGPGGCFEGG
AGNCSSRGGRASDHPPGGREFFFDRLHPGVFAYVLNYYRTGKLHCPADVCGPLFEEELAFWGI DETDVEPCWMTYRQHRDA
EEALDIFETPDLIGDPPGDDDLAAKRLGIEDAAGLGPGDKSGRWRRLQPRMWALFEDPYSSRAARFIAFASLFFILVSI
TTFCLETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVVVFTEFLVRI VFSPNKLFIKNLLNIIDFVAIL
PFYLEVGLSGLSSKAAKDLGFLRVRFVRIILRI FKLTRHPVGLRVLGHTRASTNEFLLLIIFLAGVLI FATMIYYAER
VGAQPNDPSEHTQFKNIPIGFWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAMVPVIVNNFGMYSLAMAK
```

QKLPRKRKKHIPPAPQASSPTFCCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPLSPPERLPPIRRSSTRD
KNRRGETCFLLTGDTYTCASDGGIRKGYEKSRLNNIAGLAGNALRLSPVTSPYNSPCPLRRSRSPISIL

The NOV1 amino acid sequence has 623 of 638 amino acid residues (97%) identical to, and 625 of 638 amino acid residues (97%) similar to, a *Rattus norvegicus* 638 amino acid residue voltage-gated potassium channel protein KV3.2 (KSHIIIA) (ptnr:SWISSPROT-ACC:P22462) (E = 0.0).

NOV1 is expressed in at least the following tissues: brain and lung. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, and/or RACE sources. In addition, NOV1 is predicted to be expressed in brain tissues because of the expression pattern of a closely related *Rattus norvegicus* K⁺ channel protein (KSHIIIA3) mRNA homolog (gb:GENBANK-ID:RATSHIIC|acc:M84203.1).

NOV1 has homology to the amino acid sequences shown in the BLASTP data listed in Table 1C.

Table 1C. BLAST results for NOV1					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16611597 gb AAL2 7272.1 AF268896_1 (AF268896)	voltage gated potassium channel Kv3.2b [Homo sapiens]	638	531/638 (83%)	531/638 (83%)	0.0
gi 116440 sp P22462 CIKE_RAT	VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KV3.2 (KSHIIIA)	638	520/638 (81%)	522/638 (81%)	0.0
gi 16611600 gb AAL2 7273.1 AF268897_1 (AF268897)	voltage gated potassium channel Kv3.2a [Homo sapiens]	613	486/593 (81%)	486/593 (81%)	0.0
gi 112304 pir A394 02	potassium channel protein IIIA form 1, shaker-type [Rattus norvegicus]	613	475/593 (80%)	477/593 (80%)	0.0
gi 285134 pir S227 03	voltage-gated potassium channel protein Rawl [Rattus norvegicus]	624	474/593 (79%)	476/593 (79%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 1D. In the ClustalW alignment of the NOV1 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and

can potentially be altered to a much broader extent without altering protein structure or function.

Table 1D. ClustalW Analysis of NOV1

- 1) Novel NOV1 (SEQ ID NO:2)
- 2) [gi|16611597|gb|AAL27272.1|AF268896_1](#) (AF268896) voltage gated potassium channel Kv3.2b [Homo sapiens] (SEQ ID NO:43)
- 3) [gi|116440|sp|P22462|CIKE_RAT](#) VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KV3.2 (KSHIIIA) (SEQ ID NO:44)
- 4) [gi|16611600|gb|AAL27273.1|AF268897_1](#) (AF268897) voltage gated potassium channel Kv3.2a [Homo sapiens] (SEQ ID NO:45)
- 5) [gi|112304|pir|A39402](#) potassium channel protein IIIA form 1, shaker-type [Rattus norvegicus] (SEQ ID NO:46)
- 6) [gi|285134|pir|S22703](#) voltage-gated potassium channel protein Raw1 [Rattus norvegicus] (SEQ ID NO:47)

		10	20	30	40	50	60	70	
NOV1		MGKIENNERVILNVGGTRHET	YRSTLKTLP	GTRLALLASSE	PGDCLTTAGDKLQPS	PPPLSPPPR	APPL		
gi 16611597		MGKIENNERVILNVGGTRHET	YRSTLKTLP	GTRLALLASSE	PGDCLTTAGDKLQPS	PPPLSPPPR	APPL		
gi 116440		MGKIENNERVILNVGGTRHET	YRSTLKTLP	GTRLALLASSE	PGDCLTTAGDKLQPS	PPPLSPPPR	APPL		
gi 16611600		MGKIENNERVILNVGGTRHET	YRSTLKTLP	GTRLALLASSE	PGDCLTTAGDKLQPS	PPPLSPPPR	APPL		
gi 112304		MGKIENNERVILNVGGTRHET	YRSTLKTLP	GTRLALLASSE	PGDCLTTAGDKLQPS	PPPLSPPPR	APPL		
gi 285134		MGKIENNERVILNVGGTRHET	YRSTLKTLP	GTRLALLASSE	PGDCLTTAGDKLQPS	PPPLSPPPR	APPL		
		80	90	100	110	120	130	140	
NOV1		SPGPGGCFEGGACNCSSRGGRAS	DHPGGGREFF	FDHRPGVFAYVLN	YYRTGKLHCPAD	VCGPLFEEELAF			
gi 16611597		SPGPGGCFEGGACNCSSRGGRAS	DHPGGGREFF	FDHRPGVFAYVLN	YYRTGKLHCPAD	VCGPLFEEELAF			
gi 116440		SPGPGGCFEGGACNCSSRGGRAS	DHPGGGREFF	FDHRPGVFAYVLN	YYRTGKLHCPAD	VCGPLFEEELAF			
gi 16611600		SPGPGGCFEGGACNCSSRGGRAS	DHPGGGREFF	FDHRPGVFAYVLN	YYRTGKLHCPAD	VCGPLFEEELAF			
gi 112304		SPGPGGCFEGGACNCSSRGGRAS	DHPGGGREFF	FDHRPGVFAYVLN	YYRTGKLHCPAD	VCGPLFEEELAF			
gi 285134		SPGPGGCFEGGACNCSSRGGRAS	DHPGGGREFF	FDHRPGVFAYVLN	YYRTGKLHCPAD	VCGPLFEEELAF			
		150	160	170	180	190	200	210	
NOV1		WGIDETDVEPCCWMTYRQHRDAE	EALDI	FETPD	LIGGDPGDD	EDLAAKRLGI	EDAAGLGGPDGKSGRWR		
gi 16611597		WGIDETDVEPCCWMTYRQHRDAE	EALDI	FETPD	LIGGDPGDD	EDLAAKRLGI	EDAAGLGGPDGKSGRWR		
gi 116440		WGIDETDVEPCCWMTYRQHRDAE	EALDI	FETPD	LIGGDPGDD	EDLAAKRLGI	EDAAGLGGPDGKSGRWR		
gi 16611600		WGIDETDVEPCCWMTYRQHRDAE	EALDI	FETPD	LIGGDPGDD	EDLAAKRLGI	EDAAGLGGPDGKSGRWR		
gi 112304		WGIDETDVEPCCWMTYRQHRDAE	EALDI	FETPD	LIGGDPGDD	EDLAAKRLGI	EDAAGLGGPDGKSGRWR		
gi 285134		WGIDETDVEPCCWMTYRQHRDAE	EALDI	FETPD	LIGGDPGDD	EDLAAKRLGI	EDAAGLGGPDGKSGRWR		
		220	230	240	250	260	270	280	
NOV1		LQPRMWALFEDPYSSRAARFIAFAS	FFILVSI	ITFCLETHEAFN	IVKNKTEP	VINGTS	SVVLQYEI	ETDF	
gi 16611597		LQPRMWALFEDPYSSRAARFIAFAS	FFILVSI	ITFCLETHEAFN	IVKNKTEP	VINGTS	SVVLQYEI	ETDF	
gi 116440		LQPRMWALFEDPYSSRAARFIAFAS	FFILVSI	ITFCLETHEAFN	IVKNKTEP	VINGTS	SVVLQYEI	ETDF	
gi 16611600		LQPRMWALFEDPYSSRAARFIAFAS	FFILVSI	ITFCLETHEAFN	IVKNKTEP	VINGTS	SVVLQYEI	ETDF	
gi 112304		LQPRMWALFEDPYSSRAARFIAFAS	FFILVSI	ITFCLETHEAFN	IVKNKTEP	VINGTS	SVVLQYEI	ETDF	
gi 285134		LQPRMWALFEDPYSSRAARFIAFAS	FFILVSI	ITFCLETHEAFN	IVKNKTEP	VINGTS	SVVLQYEI	ETDF	
		290	300	310	320	330	340	350	
NOV1		ALTYVEGVCVVWFTFEFLVRI	VSPN	KLEFI	KNLLNI	IDFVAIL	LPFYLEVGLSGLSSKA	AKDVLGFLRVV	
gi 16611597		ALTYVEGVCVVWFTFEFLVRI	VSPN	KLEFI	KNLLNI	IDFVAIL	LPFYLEVGLSGLSSKA	AKDVLGFLRVV	
gi 116440		ALTYVEGVCVVWFTFEFLVRI	VSPN	KLEFI	KNLLNI	IDFVAIL	LPFYLEVGLSGLSSKA	AKDVLGFLRVV	
gi 16611600		ALTYVEGVCVVWFTFEFLVRI	VSPN	KLEFI	KNLLNI	IDFVAIL	LPFYLEVGLSGLSSKA	AKDVLGFLRVV	
gi 112304		ALTYVEGVCVVWFTFEFLVRI	VSPN	KLEFI	KNLLNI	IDFVAIL	LPFYLEVGLSGLSSKA	AKDVLGFLRVV	
gi 285134		ALTYVEGVCVVWFTFEFLVRI	VSPN	KLEFI	KNLLNI	IDFVAIL	LPFYLEVGLSGLSSKA	AKDVLGFLRVV	
		360	370	380	390	400	410	420	
NOV1		RFVRIILRI	FKLTRHFVGLRVLGHT	TLRAST	NEFLLLI	I	FLALGVLI	FATMIYYAERVGAQNP	DP
gi 16611597		RFVRIILRI	FKLTRHFVGLRVLGHT	TLRAST	NEFLLLI	I	FLALGVLI	FATMIYYAERVGAQNP	DP
gi 116440		RFVRIILRI	FKLTRHFVGLRVLGHT	TLRAST	NEFLLLI	I	FLALGVLI	FATMIYYAERVGAQNP	DP
gi 16611600		RFVRIILRI	FKLTRHFVGLRVLGHT	TLRAST	NEFLLLI	I	FLALGVLI	FATMIYYAERVGAQNP	DP
gi 112304		RFVRIILRI	FKLTRHFVGLRVLGHT	TLRAST	NEFLLLI	I	FLALGVLI	FATMIYYAERVGAQNP	DP
gi 285134		RFVRIILRI	FKLTRHFVGLRVLGHT	TLRAST	NEFLLLI	I	FLALGVLI	FATMIYYAERVGAQNP	DP


```

      430      440      450      460      470      480      490
NOV1  FKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAAMPVPVIVNNFGMYYSLAMAKQKLP
gi|16611597| FKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAAMPVPVIVNNFGMYYSLAMAKQKLP
gi|116440| FKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAAMPVPVIVNNFGMYYSLAMAKQKLP
gi|16611600| FKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAAMPVPVIVNNFGMYYSLAMAKQKLP
gi|112304| FKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAAMPVPVIVNNFGMYYSLAMAKQKLP
gi|285134| FKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAAMPVPVIVNNFGMYYSLAMAKQKLP

      500      510      520      530      540      550      560
NOV1  RKRRKKHIPPAPQASSPTFCCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPPLSPPERLPI
gi|16611597| RKRRKKHIPPAPQASSPTFCCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPPLSPPERLPI
gi|116440| RKRRKKHIPPAPQASSPTFCCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPPLSPPERLPI
gi|16611600| RKRRKKHIPPAPQASSPTFCCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPPLSPPERLPI
gi|112304| RKRRKKHIPPAPQASSPTFCCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPPLSPPERLPI
gi|285134| RKRRKKHIPPAPQASSPTFCCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPPLSPPERLPI

      570      580      590      600      610      620      630
NOV1  RRSSTRDKNRRGETCFLLTTGDTYTCASDGGIRKGYEKSRSLNNIAGLACNALRLSPVTSPPYNSPCPLRRS
gi|16611597| RRSSTRDKNRRGETCFLLTTGDTYTCASDGGIRKGYEKSRSLNNIAGLACNALRLSPVTSPPYNSPCPLRRS
gi|116440| RRSSTRDKNRRGETCFLLTTGDTYTCASDGGIRKGYEKSRSLNNIAGLACNALRLSPVTSPPYNSPCPLRRS
gi|16611600| RRSSTRDKNRRGETCFLLTTGDTYTCASDGGIRKGYEKSRSLNNIAGLACNALRLSPVTSPPYNSPCPLRRS
gi|112304| RRSSTRDKNRRGETCFLLTTGDTYTCASDGGIRKGYEKSRSLNNIAGLACNALRLSPVTSPPYNSPCPLRRS
gi|285134| RRSSTRDKNRRGETCFLLTTGDTYTCASDGGIRKGYEKSRSLNNIAGLACNALRLSPVTSPPYNSPCPLRRS

      .....|...
NOV1  RSIPIPSIL
gi|16611597| RSIPIPSIL
gi|116440| RSIPIPSIL
gi|16611600| -----
gi|112304| -----
gi|285134| -----

```

The presence of identifiable domains in NOV1, as well as all other NOVX proteins, was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (<http://www.ebi.ac.uk/interpro>). DOMAIN results for NOV1, as disclosed in Tables 1E – 1G, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Tables 1E, 1F, 1G and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading or by the sign (!) and “strong” semi-conserved residues are indicated by grey shading or by the sign (+). The “strong” group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Tables 1E – 1G lists the domain description from DOMAIN analysis results against NOV1. This indicates that the NOV1 sequence has properties similar to those of other proteins known to contain these domains.

Table 1E. Domain Analysis of NOV1

gnl|Pfam|pfam02214, K_tetra, K+ channel tetramerisation domain. The N-terminal, cytoplasmic tetramerization domain (T1) of voltage-gated K+ channels encodes molecular determinants for subfamily-specific assembly of alpha-subunits into functional tetrameric channels. It is distantly related to the BTB/POZ domain pfam00651. (SEQ ID NO:48)
Length = 99 residues, 99.0% aligned
Score = 107 bits (267), Expect = 2e-24

NOV1:	9	RVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTAGDKLQSPPPPLSPPPRAP	68
		+	
02214:	1	RVRLNVGGKRFETSKSTLTRFPDTRLGRL-----	29
NOV1:	69	PLSPGPGGCFEGGAGNCSSRGGGRASDHPPGGGREFFDRHPGVFAYVLNYYRTG-KLHCPA	127
		+ ++ +	
02214:	30	-----LECRDCDFYDDATG--EYFFDRSPKHFEHILNFYRTGGKLHRPE	71
NOV1:	128	DVCGPLFEEELAFWGIDET DVEPCWMM	154
		+ + + +	
02214:	72	EVCLESFLEELEFYGLDELAIELCCED	98

Table 1F. Domain Analysis of NOV1

gnl|Pfam|pfam00520, ion_trans, Ion transport protein. This family contains Sodium, Potassium, Calcium ion channels. This family is 6 transmembrane helices in which the last two helices flank a loop which determines ion selectivity. In some sub-families (e.g. Na channels) the domain is repeated four times, whereas in others (e.g. K channels) the protein forms as a tetramer in the membrane. A bacterial structure of the protein is known for the last two helices but is not the Pfam family due to it lacking the first four (SEQ ID NO:49)
Length = 191 residues, 96.9% aligned
Score = 89.7 bits (221), Expect = 4e-19

NOV1:	282	LTYVEGVCVVWFTFEFLVRIVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAAK	341
		++ ++ + +++++ + + + ++ +	
00520:	2	LEILDYVFTVIFTLEMLLKFIALGFKLYLRSPWNILDFLIVLPSLIDLILFLSGGGSV-	60
NOV1:	342	DVLGFLRVVRVFRILRIFKLTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLI FATMIY	401
		+ + + + + + ++ ++ ++	
00520:	61	----LRLRLRLRLRLRLRLEGLRTLQSLGRSLKSLN-LLLLLLLLLFI FAIIGVQLF	115
NOV1:	402	YAERVGAQPNDPSASEHTQFKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAG	461
		+ ++ + + + + ++	
00520:	116	GGEFNKCCDGVNPNINGNSNFDSFGGEAFYWLFRTLTTEGWGDIMPDTLDAP-----	165
NOV1:	462	VLTIAMPVVPVIVNPNFGMYYSLAM	484
		+ +	
00520:	166	--VLGKIFFVIFIILGGLLLNL	186

Table 1G. Domain Analysis of NOV1

gnl|Smart|smart00225, BTB, Broad-Complex, Tramtrack and Bric a brac; Domain in Broad-Complex, Tramtrack and Bric a brac. Also known as POZ (poxvirus and zinc finger) domain. Known to be a protein-protein interaction motif found at the N-termini of several C2H2-type transcription factors as well as Shaw-type potassium channels. Known structure reveals a tightly intertwined dimer formed via interactions between N-terminal strand and helix structures. However in a subset of BTB/POZ domains, these two secondary structures appear to be missing. Be aware SMART predicts BTB/POZ domains without the beta1- and alpha1-secondary structures. (SEQ ID NO:50)
Length = 96 residues, 99.0% aligned
Score = 49.7 bits (117), Expect = 5e-07

```
NOV1: 10  VILNVGGTRHETYRSTLTKLPGTRLALLASSEPPGDCLTTAGDKLQSPSPPLSPPPRAPP 69
      | |||| +   +++ |           ||
00225: 2  VTLNVGGKKFHAHKAVLAHSPYFKALF----- 29

NOV1: 70  LSPGPGGCFEGGAGNCSSRGGRASDHPPGGGREFFFDHRHGVFAYVLNYYRTGKLHCPADV 129
      | + || | + || | | +|| + ||| | +
00225: 30  -----SSDFKESDKS---EIYLFDVSPEDFRALLNFLYTGKLDIP-EE 68

NOV1: 130 CGPLFEEELAFWGIDET DVEPCCWMTYRQ 158
      | + | || | +
00225: 69  NVEELLELADYLQIPG-LVELCEEFLKN 96
```

Cation channels are transport proteins responsible for the movement of cations through the membrane. This family contains sodium, potassium and calcium ion channels. These proteins contain 6 transmembrane helices in which the last two helices flank a loop which determines ion selectivity. In some sub-families (e.g. Na channels) the domain is repeated four times, whereas in others (e.g. K channels) the protein forms as a tetramer in the membrane (IPR000636). The N-terminal, cytoplasmic tetramerization domain (T1) of voltage-gated K⁺ channels encodes molecular determinants for subfamily-specific assembly of alpha-subunits into functional tetrameric channels. It is distantly related to the BTB/POZ domain PFAM PF00651 (IPR003131).

Potassium channels represent a complex class of voltage-gated ion channels. These channels maintain membrane potential, regulate cell volume, and modulate electrical excitability in neurons. The delayed rectifier function of potassium channels allows nerve cells to efficiently repolarize following an action potential. NOV1 is a human ortholog of a rat voltage gated potassium channel protein, one of a large family of proteins which play crucial roles in many tissues, particularly brain and heart. Voltage gated potassium channel proteins are currently targeted for pharmaceutical intervention. These treatments are indicated for neurological disorders such as epilepsy, and cardiac disorders involving arrhythmias.

Electrophysiological studies have shown that a number of different types of potassium (K) channel currents exist in mammalian neurons. Among them are the voltage-gated K

channel-currents which have been classified as fast-inactivating A-type currents (KA) and slowly inactivating delayed-rectifier type currents (KDR). Two major molecular superfamilies of K channel have been identified; the KIR superfamily and the Shaker-related superfamily with a number of different pore-forming alpha-subunits in each superfamily. Within the Shaker-related superfamily are the KV family, comprising of at least 18 different alpha-subunits that almost certainly underlie classically defined KA and KDR currents. However, the relationship between each of these cloned alpha-subunits and native voltage-gated K currents remains, for the most part, to be established. Classical pharmacological blockers of voltage-gated K channels such as tetraethylammonium ions (TEA), 4-aminopyridine (4-AP), and certain toxins lack selectivity between different native channel currents and between different cloned K channel currents. A number of other agents block neuronal voltage-gated K channels. All of these compounds are used primarily for other actions they possess. They include organic calcium (Ca) channel blockers, divalent and trivalent metal ions and certain calcium signalling agents such as caffeine. A number of clinically active tricyclic compounds such as imipramine, amitriptyline, and chlorpromazine are also potent inhibitors of neuronal voltage-gated K channels. These compounds are weak bases and it appears that their uncharged form is required for activity. These compounds may provide a useful starting point for the rational design of novel selective K channel blocking agents (Mathie et al., Voltage-activated potassium channels in mammalian neurons and their block by novel pharmacological agents. *Gen Pharmacol* 30(1):13-24, 1998).

Subfamilies of voltage-activated K⁺ channels (Kv1-4) contribute to controlling neuron excitability and the underlying functional parameters. Genes encoding the multiple alpha subunits from each of these protein groups have been cloned, expressed and the resultant distinct K⁺ currents characterized. The predicted amino acid sequences showed that each alpha subunit contains six putative membrane-spanning alpha-helical segments (S1-6), with one (S4) being deemed responsible for the channels' voltage sensing. Additionally, there is an H5 region, of incompletely defined structure, that traverses the membrane and forms the ion pore; residues therein responsible for K⁺ selectivity have been identified. Susceptibility of certain K⁺ currents produced by the Shaker-related subfamily (Kv1) to inhibition by alpha-dendrotoxin has allowed purification of authentic K⁺ channels from mammalian brain. These are large (M(r) approximately 400 kD), octomeric sialoglycoproteins composed of alpha and beta subunits in a stoichiometry of (alpha)₄(beta)₄, with subtypes being created by combinations of subunit isoforms. Subsequent cloning of the genes for beta 1, beta 2 and beta 3 subunits revealed novel sequences for these hydrophilic proteins that are postulated to be

associated with the alpha subunits on the inner side of the membrane. Coexpression of beta 1 and Kv1.4 subunits demonstrated that this auxiliary beta protein accelerates the inactivation of the K⁺ current, a striking effect mediated by an N-terminal moiety (Dolly and Parcej, Molecular properties of voltage-gated K⁺ channels. *J Bioenerg Biomembr* 28(3):231-53, 1996).

Mice lacking the voltage-gated potassium channel alpha subunit, K(V)1.1, display frequent spontaneous seizures throughout adult life. These data indicate that loss of K(V)1.1 from its normal localization in axons and terminals of the CA3 region results in increased excitability in the CA3 recurrent axon collateral system, perhaps contributing to the limbic and tonic-clonic components of the observed epileptic phenotype (Smart et al., Deletion of the K(V)1.1 potassium channel causes epilepsy in mice. *Neuron* 20(4):809-19, 1998). Other studies indicate that Kv1.1 plays an important role in nociceptive and antinociceptive signaling pathways (Clark and Tempel, Hyperalgesia in mice lacking the Kv1.1 potassium channel gene. *Neurosci Lett* 251(2):121-4, 1998).

Histamine-containing neurons of the tuberomammillary nucleus project to the hippocampal formation to innervate H1 and H2 receptors on both principal and inhibitory interneurons. Studies show that H2 receptor activation negatively modulates outward currents through Kv3.2-containing potassium channels by a mechanism involving PKA phosphorylation in inhibitory interneurons (Atzori et al., H2 histamine receptor-phosphorylation of Kv3.2 modulates interneuron fast spiking. *Nat Neurosci* 3(8):791-8, 2000).

Classical cardiac delayed rectifier currents activate at least two orders of magnitude slower than delayed rectifier currents in nerve and skeletal muscle tissue. It has recently become evident that many cardiac tissues express delayed rectifier currents with kinetics similar to those of nerve and muscle (Nattel et al., Cardiac ultrarapid delayed rectifiers: a novel potassium current family of functional similarity and molecular diversity. *Cell Physiol Biochem* 9(4-5):217-26, 1999).

The above defined information for NOV1 suggests that this potassium channel-like protein may function as a member of the potassium channel protein family. Therefore, the NOV1 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the NOV1 compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia,

leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy and/or ARDS. The NOV1 nucleic acid encoding potassium channel-like protein, and the potassium channel-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV2

A disclosed NOV2 nucleic acid of 1227 nucleotides (also referred to as CG50293-01) encoding a novel Galanin receptor type 1 (GALR1)-like protein is shown in Table 2A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 37-39 and ending with a TAA codon at nucleotides 1225-1227. A putative untranslated region upstream from the initiation codon is underlined in Table 2A, and the start and stop codons are in bold letters.

Table 2A. NOV2 nucleotide sequence (SEQ ID NO:3).

CTGGCAGCTGCCTTTGTCAGACTCTAACTCCAGCAGCATGAATGTGTCCTTTGCTCACCTCCACTTGTCCGGA
GGGTACCTGCCCTCTGATTCCAGGACTGGAGAACCATCATCCCGGCTCTCTGGTGGCTGTCTGCCTGGTG
GGCTTCGTGGGAAACCTGTGTGTGATTGGCATCCTCCTTCACAATGCTTGGAAAGGAAAGCCATCCATGATC
CACTCCCTGATTCTGAATCTCAGCCTGGCTGATCTCTCCCTCCTGCTGTTTTCTGCACCTATCCGAGCTACG
GCGTACTCCAAAAGTGTGTTGGGATCTAGGCTGGTTGTCTGCAAGTCCTCTGACTGGTTATCCACACATGC
ATGGCAGCCCAAGAGCCTGACAATCGTTGTGGTGGCCAAAGTATGCTTCATGTATGCAAGTGACCCAGCCAAG
CAAGTGAGTATCCACAACCTACACCATCTGGTCACTGCTGGTGGCCATCTGGACTGTGGCTAGCCTGTTACCC
CTGCCGGAATGGTTCTTTAGCACCATCAGGCATCATGAAGGTGTGGAAATGTGCCTCGTGGATGTACCAGCT
GTGGCTGAAGAGTTTATGTCGATGTTTGGTAAGCTCTACCCACTCCTGGCATTGTCGCTTCCATTATTTTTT
GCCAGCTTTTATTTCTGGAGAGCTTATGACCAATGTAAAAACGAGGAAGTAAGACTCAAAATCTAGAAAC
CAGATACGCTCAAAGCAAGTCACAGTGATGCTGCTGAGCATTGCCATCATCTCTGCTCTCTTGTGGCTCCCC
GAATGGGTAGCTTGGCTGTGGTATGGCATCTGAAGGCTGCAGGCCCGGCCCCACCACAAGTTTCATAGCC
CTGTCTCAAGTCTTGATGTTTTCCATCTCTTCAGCAAATCCTCTCATTCTTCTGTGATGTCGGAAGAGTTC
AGGGAAGGCTTGAAAGGTGTATGGAATGGATGATAACCAAAAAACCTCCAAGTCTCAGAGTCTCAGGAA
ACACCAGCTGGCAACTCAGAGGCTCTCCTGACAAGGTTCCATCTCCAGAATCCCAGCATCCATACCAGAA
AAAGAGAAACCCAGCTCTCCCTCCTCTGGCAAAGGGAAAAGTGAAGGCAGAGATCCCATCTCTCTGAC
GTAGAGCAGTTTTGGCATGAGAGGGACACAGTCCCTTCTGTACAGGACAATGACCTATCCCCCTGGGAACA
TAA

The disclosed NOV2 nucleic acid sequence, localized to chromosome 5, has no homology to any known nucleic acid sequence.

A NOV2 polypeptide (SEQ ID NO:4) encoded by SEQ ID NO:3 has 396 amino acid residues and is presented using the one-letter code in Table 2B. Signal P, Psort and/or Hydropathy results predict that NOV2 contains a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a NOV2 peptide is between amino acids 41 and 42, at: VGN-LC.

Table 2B. Encoded NOV2 protein sequence (SEQ ID NO:4).

MNVSFALHFFAGGYLPSDSQDWRTIIIPALLVAVCLVGVGNLCVIGILHNAWKGPMSIHLIHLNLADL

SLLLFSAPIRATAYSKSVWDLGWVCKSSDWFIHTCMAAKSLTIVVAVKVCIFYASDPKQVSIHNYTIWSV
LVAIWTVASLLPLPEWFFSTIRHHEGVEMCLVDVPAVAEEFMSMFGKLYPLLAFLPLFFASFYFWRAYDQC
KKRGTKTQNLNRNQIRSKQVTVMLLSIAIISALLWLPEWVAWLWVWHLKAAGPAPPQGFIALSQVLMFSISSA
NPLIFLVMSEEFREGLKGVWKWMITKKPPTVSESQETPAGNSEGLPKVPSPEPASIPKEKEKPSPPSSGKG
KTEKAETIPILPDVEQFWHERDTPVPSVDNDPIPLGT

The NOV2 amino acid sequence has 80 of 289 amino acid residues (27%) identical to, and 135 of 289 amino acid residues (46%) similar to, a *Homo sapiens* 349 amino acid residue Galanin receptor type 1 (GAL1-R) protein (ptnr:SWISSPROT-ACC:P47211) ($E = 5.0e^{-21}$).

5 The disclosed NOV2 is expressed in at least the following tissues: brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

10 Possible small nucleotide polymorphisms (SNPs) found for NOV2 are listed in Tables 2C and 2D. Depth represents the number of clones covering the region of the SNP. The putative allele frequency (PAF) is the fraction of these clones containing the SNP. A dash, when shown, means that a base is not present. The sign ">" means "is changed to."

Table 2C: SNPs			
Consensus Position	Depth	Base Change	PAF
121	38	C > T	N/A
395	33	G > C	N/A
786	34	C > G	N/A

Table 2D: SNPs				
Variant	Nucleotide Position	Base Change	Amino Acid Position	Base Change
13373926	116	C > T	27	Pro > Leu
13373927	388	G > C	118	Val > Leu
13373928	778	C > G	248	Leu > Val

15 NOV2 has homology to the amino acid sequences shown in the BLASTP data listed in Table 2E.

Table 2E. BLAST results for NOV2					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 6016095 sp O88854 GALS_MOUSE	GALANIN RECEPTOR TYPE 2 (GAL2-R) (GALR2) [Mus musculus]	371	64/314 (20%)	112/314 (35%)	1e-08

gi 4503905 ref NP_03848.1 (NM_003857)	galanin receptor 2 [Homo sapiens]	387	66/316 (20%)	114/316 (35%)	4e-08
gi 9506709 ref NP_062045.1 (NM_019172)	galanin receptor 2 [Rattus norvegicus]	372	64/315 (20%)	111/315 (34%)	5e-08
gi 6753942 ref NP_034384.1 (NM_010254)	galanin receptor 2 [Mus musculus]	371	63/313 (20%)	111/313 (35%)	3e-07
gi 9506711 ref NP_062046.1 (NM_019173)	galanin receptor 3 [Rattus norvegicus]	370	40/160 (25%)	68/160 (42%)	3e-05

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 2F.

Table 2F. ClustalW Analysis of NOV2

- 1) NOV2 (SEQ ID NO:4)
- 2) gi|6016095|sp|O88854|GALS_MOUSE GALANIN RECEPTOR TYPE 2 (GAL2-R) (GALR2) [Mus musculus] (SEQ ID NO:51)
- 2) gi|4503905|ref|NP_003848.1| (NM_003857) galanin receptor 2 [Homo sapiens] (SEQ ID NO:52)
- 3) gi|9506709|ref|NP_062045.1| (NM_019172) galanin receptor 2 [Rattus norvegicus] (SEQ ID NO:53)
- 4) gi|6753942|ref|NP_034384.1| (NM_010254) galanin receptor 2 [Mus musculus] (SEQ ID NO:54)
- 5) gi|9506711|ref|NP_062046.1| (NM_019173) galanin receptor 3 [Rattus norvegicus] (SEQ ID NO:55)

	10	20	30	40	50	60	70
NOV2	MNVSF	AHLHFAGGYLPS	DSQWR	TTPALLVAV	CLVGFVGNL	CVIGILLHN	AWK---
gi 6016095	MNGSD	SGCAEDSSCEGG	GGWOPEAV	LVPLFFALI	FLVGA	VGNALVLA	VLLRGG---
gi 4503905	MNVSG	CFGAGNASCAGG	GGWHEPAV	TVPLFFALI	FLVGT	VGNLVL	LAVLLRGG---
gi 9506709	MNGSC	SGCAENTSCGCS	GGWOPEAV	LVPLFFALI	FLVGT	VGNALVLA	VLLRGG---
gi 6753942	MNGSD	SGCAEDSSCEGG	GGWOPEAV	LVPLFFALI	FLVGA	VGNALVLA	VLLRGG---
gi 9506711	MADIQ	NISLSPGSVC	-----	AVAVP	IFALI	FLVGMVGNL	VLA
	80	90	100	110	120	130	140
NOV2	NLSAD	LSLLFSA	FIRATAY	SKSVMDL	GWFKSSD	WFHTOM	AAKSLTE
gi 6016095	NLGVAD	LCFILCCV	PFQATIT	YTLDDW	VFGSLLC	KAVHFLI	FLTMH
gi 4503905	NLGVAD	LCFILCCV	PFQATIT	YTLDDW	VFGSLLC	KAVHFLI	FLTMH
gi 9506709	NLGVAD	LCFILCCV	PFQATIT	YTLDDW	VFGSLLC	KAVHFLI	FLTMH
gi 6753942	NLGVAD	LCFILCCV	PFQATIT	YTLDDW	VFGSLLC	KAVHFLI	FLTMH
gi 9506711	NLAVAD	LCFILCCV	PFQATIT	YTLDAW	FGAFVCK	TVHLLI	FLTMH
	150	160	170	180	190	200	210
NOV2	THN-Y	TIWSVL	VAIWT	NASLL	PLFEW	FESTIR	HHEGVEM
gi 6016095	LRTPR	NALAAI	GLIWL	LALLF	SGPYLS	YYROS	QLANLT
gi 4503905	LRTPR	NALAAI	GLIWL	LALLF	SGPYLS	YYROS	QLANLT
gi 9506709	LRTPR	NALAAI	GLIWL	LALLF	SGPYLS	YYROS	QLANLT
gi 6753942	LRTPR	NALAAI	GLIWL	LALLF	SGPYLS	YYROS	QLANLT
gi 9506711	LRTPR	NALAAV	GLIWL	LALLF	SGPYLS	YYGTV	RYGAL
	220	230	240	250	260	270	280
NOV2	PLFFAS	FVWRA	YDOCK	KRGTK	TONTLR	NCIRSK	CVTVML
gi 6016095	TYART	LHYLWR	IVDPV	-----	AAGSG	SQRAKR	KVTRMI
gi 4503905	TYART	LHYLWR	IVDPV	-----	AAGSG	SQRAKR	KVTRMI
gi 9506709	TYART	LHYLWR	IVDPV	-----	AAGSG	SQRAKR	KVTRMI
gi 6753942	TYART	LHYLWR	IVDPV	-----	AAGSG	SQRAKR	KVTRMI
gi 9506711	AYCRT	LCELWA	AVG	PAGA	-----	AAAEARR	ATG
	290	300	310	320	330	340	350
NOV2	PQGFIA	LSQVLM	FESIS	SANLE	LILVMS	EBPRE	ELG


```

gi|6016095|TYALRILSHLVSYANSCVNPVYALVSKHFRKGFRKIC-AGLLRRAPRRASGRVCILAPGNHSGCMLEPE
gi|4503905|TYALRILSHLVSYANSCVNPVYALVSKHFRKGFRKIC-AGLLGRAPGRASGRVCAAPRGTHSGSVLERE
gi|9506709|TYALRILSHLVSYANSCVNPVYALVSKHFRKGFRKIC-AGLLRPAPRRASGRVSILAPGNHSGCMLEPE
gi|6753942|TYALRILSHLVSYANSCVNPVYALVSKHFRKGFRKIC-AGLLRPAPRRASGRVCILAPGNHSGCMLEPE
gi|9506711|TYACRLASHCLAYANSCVNPVYSLASHFRFRFRRLWPCRRRRHRRHRRHRRALRRVQPPASSCPAGYEG

          360      370      380      390      400
NOV2      PESPASIPPEKEKPSSESSGKGKTEKAEIPILPDVEQFWHERDTVPSVQNDPIPLGT
gi|6016095|SDLTQVSEAAAGPLVP-AP---ALPNCITLSR-----TLDPA---
gi|4503905|SSDLLHMEAAAGALRCPG---ASQPCILEPCPGPSWQGPAGDSILTVDA---
gi|9506709|SDLTQVSEAAAGPLVP-PP---ALPNCIASSR-----TLDPA---
gi|6753942|SDLTQVKGCSRAPRFRTR---TSQTXITLSR-----TLDPA---
gi|9506711|DARPRGWSMEPRGDALRGG---GETRLTLSR-----GEQ---

```

Table 2G list the domain description from DOMAIN analysis results against NOV2. This indicates that the NOV2 sequence has properties similar to those of other proteins known to contain these domains.

Table 2G Domain Analysis of NOV2

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:56)
Length = 254 residues, 80.7% aligned
Score = 46.6 bits (109), Expect = 3e-06

```

NOV2: 91 WDLGWVCKSSDWFIIHTCMAAKSLTIVVAKVCFMYASDPKQVSIHNYTI-WSVLVAIW 149
      | | +|| | | + ++ ++ | + | +++ +|
00001: 50 WVFGDALCKLVGALFVVNGYASILLTATSIDRYLAIVHPLRYRRIRTPRRAKVLILLVW 109

NOV2: 150 TVASLLPLPEWFFSTIRHHEGVEMCLVDVPAVAEEFMSMFGLYPLLAFLPLFFASFYF 209
      +| || || || +| | + + | + | ||| +
00001: 110 VLALLLSLPLLFSLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVPLLVILVCY 169

NOV2: 210 WRAYDQCKKRGTK-TQNLNRNQIRSKQVTVMLLSIAIISALLWLPEWVAWLWVWHLKAAGP 268
      | | +|| | ++ ||| + ++ | ||| + |
00001: 170 TRILRTLKRARSQRSLSKRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLLLSLCLLSIW 229

NOV2: 269 APPQGFIALSQVLMFSSISANPLIF 293
      + ++ | + | ||| +
00001: 230 RVLPTALLITLWLAYVNSCLNPIIY 254

```

Human and rat GALR1 galanin receptor cDNA clones have previously been isolated using expression cloning. Human GALR1 cDNA in hybridization screening has been used to isolate the gene encoding GALR1 in both human (GALNR) and mouse (Galnr). The gene spans approximately 15-20 kb in both species; its structural organization is conserved and is unique among G-protein-coupled receptors. The coding sequence is contained on three exons, with exon 1 encoding the N-terminal end of the receptor and the first five transmembrane domains. Exon 2 encodes the third intracellular loop, while exon 3 encodes the remainder of the receptor, from transmembrane domain 6 to the C-terminus of the receptor protein. The mouse and human GALR1 receptor proteins are 348 and 349 amino acids long, respectively, and display 93% identity at the amino acid level. The mouse Galnr gene has been localized to Chromosome 18E4, homoeologous with the previously reported localization of the human

any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1	CG50249_01	1	2	Potassium Channel-like
2	CG50293_01	3	4	Galanin Receptor Type 1 (GAL1-R) (GALR1)-like
3	CG50237_01	5	6	P2Y Purinoceptor-1-like
4a	CG50255_01	7	8	LOMP-like
4b	CG50255-02	9	10	LOMP-like
5	16467945_0_88_da1	11	12	Epidermal Growth Factor-like
6a	CG50239_01	13	14	Hyaluronan Mediated Motility Receptor-like
6b	CG50239-02	15	16	Hyaluronan Mediated Motility Receptor-like
6c	CG50239-03	17	18	Hyaluronan Mediated Motility Receptor-like
7	AC019355.3	19	20	Serpin-like
8a	CG50309_01	21	22	B7 family-like
8b	CG50309-02	23	24	B7 family-like
8c	CG50309_03	25	26	B7 family-like
8d	CG50309-04	27	28	B7 family-like
8e	CG50309_05	29	30	B7 family-like
8f	170403925	31	32	B7 family-like
8g	169376006	33	34	B7 family-like
9a	cg-140509446	35	36	Acyl CoA Dehydrogenase-like
9b	CG55900-02	37	38	Acyl CoA Dehydrogenase-like
9c	CG55900-03	39	40	Acyl CoA Dehydrogenase-like
9d	CG55900-04	41	42	Acyl CoA Dehydrogenase-like

5 NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

10 NOV1 is homologous to a Potassium Channel-like family of proteins. Thus, the NOV1 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, for example; Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, systemic lupus erythematosus,

GALNR gene to 18q23 in the same syntenic group as the genes encoding nuclear factor of activated T-cells, cytoplasmic 1, and myelin basic protein (Jacoby et al., The neuropeptide galanin elicits a range of biological effects by interaction with specific G-protein-coupled receptors *Genomics* 45(3):496-508, 1997).

5 This conservation of structural organization is indicative of a common evolutionary origin for GALNR2 and GALNR3. The exon:intron organization of the gene encoding GALR1 (GALNR1) is different from that of GALNR2 and GALNR3, with exon 1 encoding the NH2-terminus to the end of transmembrane domain 5, exon 2 encoding the third
10 intracellular loop, and exon 3 encoding the remainder of the receptor, from transmembrane domain 6 to the COOH-terminus. The structural organization of GALNR1 suggests convergent evolution for this gene and represents a structural organization that is unique among genes encoding G-protein-coupled receptors (Iismaa et al., Human galanin receptor subtypes GALR1, GALR2, and GALR3 are encoded by separate genes that are located on human chromosomes 18q23, 17q25.3, and 22q13.1, respectively *Ann N Y Acad Sci* 863:56-
15 63, 1998).

 Studies suggest that galanin receptors mediate via different Gi/Go-proteins the inhibition of adenylyl cyclase, opening of K⁺-channels and closure of Ca²⁺-channels. Galanin inhibits secretion of insulin, acetylcholine, serotonin and noradrenaline, while it stimulates prolactin and growth hormone release. Determination of structural components of galanin
20 receptors required for binding of the peptide ligand as carried out recently will facilitate the screening and design of molecules specifically acting on galaninergic systems with therapeutic potential in Alzheimer's disease, feeding disorders, pain and depression (Kask et al., Galanin, a neuroendocrine peptide with a multitude of functions, binds to and acts on specific G-protein coupled receptors *Life Sci* 60(18):1523-33, 1997).

25 Although studies have shown that galanin peptide mRNA levels do not change during withdrawal, it is not known whether galanin receptor levels are regulated following opiate withdrawal. More recent studies demonstrate that galanin binding in the LC is upregulated by chronic-intermittent morphine administration or by precipitated withdrawal, but not by acute morphine treatment, suggesting that increased activity in the LC may be able to regulate
30 galanin binding sites. Moreover, the increase in galanin binding sites seems to be caused by increased transcription or stabilization of the galanin receptor 1 (GalR1) gene, because there is a dramatic increase in mRNA levels following withdrawal in the LC. It is, therefore, possible that the increase in GalR1 could be an adaptive mechanism that leads to regulation of cAMP levels and possibly firing rate of LC neurons (Zachariou et al., The neuropeptide galanin and

its receptors are expressed in the locus coeruleus (LC), a brain area associated with drug dependence and withdrawal Neuropsychopharmacology 23(2):127-37, 2000).

Other studies have indicated that exogenously administered galanin may stimulate ingestion, and endogenous galanin may have an affect on feeding and body weight, These studies suggest the therapeutic potential of non-peptide galanin receptor antagonists for the treatment of appetite disorders (Crawley et al., Galanin inhibits food consumption in satiated rats. Neuropeptides 33(5):369-75, 1999).

The above defined information for NOV2 suggests that the NOV2 protein may function as a member of a family of novel Galanin receptor type 1 (GALR1)-like proteins. Therefore, the NOV2 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the NOV2 compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain and/or neuroprotection. The NOV2 nucleic acid encoding Galanin receptor type 1 (GALR1)-like proteins, and the Galanin receptor type 1 (GALR1)-like proteins of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV3

A disclosed NOV3 nucleic acid of 1560 nucleotides (also referred to as CG50237-01) encoding a novel P2Y purinoceptor 1- like protein is shown in Table 3A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 353-355 and ending with a TGA codon at nucleotides 1364-1366. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 3A, and the start and stop codons are in bold letters.

Table 3A. NOV3 Nucleotide Sequence (SEQ ID NO:5)

CTAGACTAGAAATTCAGCGACCAACTAGGCTGCACAGGCACGCTGGGGCGCATGTCCGCCTCGCCGGGGCTGCCAGAATCT
TGAATCCCAATCCGTGAGGTTCTGGGTGTGCTGGCATCAGGACAGCGGTCCACGAACGGTGTGTACCCAAATATTGA
CATCGTGCAGCTAGCCTCAAACAATCACAGCTACTTCCAATTTCAGAGAAAAAGGCTAAAAATGGTAATCCTGATGA
AAATCAACAAAATACACATGAAGAGACAGCACTGAGAGCGAGTTACTGCTCATTGATTTCATATTGCCAACTGAACCTC
CTGTGTTTCTTGAAGATGAAAGGAGACAACCATGAATGAGCCACTAGACTATTTAGCAAATGCTTCTGATTTCCCGAT
TATGCAGCTGCTTTTGGAAATTGCATGATGAAAACATCCCACTCAAGATGCACACTCCCTGTTATTTATGGCAITAT
CTTCCTCGTGGGATTTCAGGCAATGCAGTAGTGATATCCACTTACATTTCAAATGAGACCTTGAAGAGCAGCACCA
TCATTATGCTGAACCTGGCCTGCACAGATCTGCTGTATCTGACCAGCCTCCCTTCTGATTCACCTACTATGCCAGTGGC
GAAAACTGGATCTTTGGAGATTTCATGTGTAAGTTTATCCGCTTCAGCTTCCATTTCAACCTGTATAGCAGCATCCTCTT
CCTCACCTGTTTCAGCATCTTCCGCTACTGTGTGATCATTACCCAATGAGCTGCTTTTCCATTCACAAAACCTCGATGTG
CAGTTGTAGCCTGTGCTGTGGTGTGGATCATTTCACTGGTAGCTGTCACTTCOGATGACCTTCTTGATCACATCAACCAAC
AGGACCAACAGATCAGCCTGTCTCGACCTCACCAGTTCGGATGAACCTAATACTATTAAGTGGTACAACCTAATTTTGAC
TGCAACTACTTTCTGCCTCCCTTGGTGTATGTGACACTTGTGTATACCAGGATATCCCACTCTGACCCATGGAGCTGC
AAACTGACAGCTGCTTAAGCAGAAAGCAGCAAGGCTAACCAATTCTGCTACTCCTTGCAATTTTACGTATGTTTTTACCC
TTCCATATCTTGAGGTCATTCGGATCGAATCTCGCTGCTTTCAATCAGTTGTTCATTGAGAATCAGATCCATGAAGC
TTACATCGTTTCTAGACCATTAGCTGCTCTGAACACCTTTGGTAACCTGTACTATATGTTGGTGGTCAAGCACAACCTTC
AGCAGGCTGTCTGCTCAACAGTGAGATGCAAGTAAGCGGGAACCTTGAGCAAGCAAGAAAAATTAGTTACTCAACCAAC
CCTTGAATATTTCATTACTTAACCAAAACAAATACTTGTGTACTTTACCTAGCATCTTAAGATGTTTCAAGATGTC
TCCCTCAATGGAACCTCTGGTAAATCTGTGTATTCAAGTAATCATGTGCCAAGCCAGGCGCAGAGCTTCTAGTCTTTT
CGTCGACGCGGCCGGAATTTAGTAGTAGTAGGCGGCCG

The disclosed NOV3 nucleic acid sequence maps to chromosome 13 and has 398 of 639 bases (62%) identical to a *Gallus gallus* G. domesticus mRNA for G protein-coupled P2 receptor (gb:GENBANK-ID:GDP2Y3|acc:X98283.1) ($E = 2.7e^{-19}$).

A disclosed NOV3 protein (SEQ ID NO:6) encoded by SEQ ID NO:5 has 337 amino acid residues, and is presented using the one-letter code in Table 3B. Signal P, Psort and/or Hydropathy results predict that NOV3 does not contain a signal peptide, and is likely to be localized to the plasma membrane with a certainty of 0.6000.

Table 3B. Encoded NOV3 protein sequence (SEQ ID NO:6).

MNEPLDYLANASDFPDYAAAFNCTDENIPLKMHYLPVIYGIIFLVGFPGNAVVIISTYIFKMRPWKSSIIIMLNACTDL
LYLTSLPFLIHYASGENWIFGDFMCKFIRFSHFNLSSILEFTCFSEIFRYCVIIHPMSCFSEIHKTRCAVAVCAVWV
SLVAVIPMTFLITSTNRTNRSACLDLTSSDELNTIKWYNLILTAITFCFLPLVIVTLCYTTIIHTLTHGLQTDSC
LKQKARRLTILLLAFYVCFPLFHLRVIRESRLISCSISNENQIHEAYIVSRPLAALNTFGNLLLYVVVSDNFQAVCSTVRCK
VSGNLEQAKKISYSNNP

The NOV3 amino acid sequence has 111 of 306 amino acid residues (36%) identical to, and 179 of 306 amino acid residues (58%) similar to; a *Homo sapiens* 373 amino acid residue (P2Y Purinoceptor 1 (ATP receptor) (P2Y1) (purinergic receptor) (ptnr:SWISSNEW-ACC:P47900) ($E = 7.5e^{-55}$).

NOV3 is expressed in at least the following tissues: brain, lung, cervix, colon, thyroid, uterus, testis, umbilical cord vein, endothelium and liver. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Genomic Clone sources, Literature sources, and/or RACE sources.

Possible small nucleotide polymorphisms (SNPs) found for NOV3 are listed in Table

Table 3C: SNPs				
Variant	Nucleotide Position	Base Change	Amino Acid Position	Base Change
13373896	366	T > C	5	Leu > Pro
13373897	640	C > T	Silent	N/A
13373898	746	T > C	132	Tyr > His
13373899	952	A > G	Silent	N/A
13373900	975	G > A	208	Cys > Tyr
13373901	1039	G > A	Silent	N/A

NOV3 has homology to the amino acid sequences shown in the BLASTP data listed in Table 3D.

Table 3D. BLAST results for NOV3					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16566323 gb AAL26480.1 AF411109_1 (AF411109)	G protein-coupled receptor [Homo sapiens]	337	328/337 (97%)	328/337 (97%)	e-168
gi 6679193 ref NP_032798.1 (NM_008772)	purinergic receptor P2Y, G-protein coupled 1; P2Y1 receptor [Mus musculus]	373	107/299 (35%)	172/299 (56%)	3e-50
gi 4505557 ref NP_002554.1 (NM_002563)	purinergic receptor P2Y, G-protein coupled, 1 [Homo sapiens]	373	106/299 (35%)	172/299 (57%)	8e-50
gi 1352695 sp P49651 P2YR_RAT	P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR) [Rattus norvegicus]	373	105/291 (36%)	170/291 (58%)	3e-49
gi 1352691 sp P48042 P2YR_BOVIN	P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR) [Bos taurus]	373	105/299 (35%)	171/299 (57%)	3e-49

5 The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 3E.

Table 3E. ClustalW Analysis of NOV3

- 1) NOV3 (SEQ ID NO:6)
- 2) gi|16566323|gb|AAL26480.1|AF411109_1 (AF411109) G protein-coupled receptor [Homo sapiens] (SEQ ID NO:57)
- 3) gi|6679193|ref|NP_032798.1| (NM_008772) purinergic receptor P2Y, G-protein coupled 1; P2Y1 receptor [Mus musculus] (SEQ ID NO:58)
- 4) gi|4505557|ref|NP_002554.1| (NM_002563) purinergic receptor P2Y, G-protein coupled, 1 [Homo sapiens] (SEQ ID NO:59)
- 5) gi|1352695|sp|P49651|P2YR_RAT P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR) [Rattus norvegicus] (SEQ ID NO:60)
- 6) gi|1352691|sp|P48042|P2YR_BOVIN P2Y PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC RECEPTOR) [Bos taurus] (SEQ ID NO:61)

		10	20	30	40	50	60	70
NOV3	
gi 16566323		MNEPLDYLANASDFPDYAAAFGN	CTDENIPLKMHYLPVITYGII	FLVGFPGNA
gi 6679193		MNEPLDYLANASDFPDYAAAFGN	CTDENIPLKMHYLPVITYGII	FLVGFPGNA
gi 4505557		MTEVLPMSVVENGTDAAFLAGLGS	WGNSTVASTAAVSSSF	OCALIKTGQFYYP	PAVYILVFI	IGFLGNS
gi 1352695		MTEVLPMSVVENGTDAAFLAGLGS	WGNSTVASTAAVSSSF	OCALIKTGQFYYP	PAVYILVFI	IGFLGNS
gi 1352691		MTEVLPMSVVENGTDAAFLAGLGS	WGNSTVASTAAVSSSF	OCALIKTGQFYYP	PAVYILVFI	IGFLGNS
		80	90	100	110	120	130	140
NOV3	
gi 16566323		VVISTYIFKMRPWKSSSTIIMLN	LACTDLYETSLP	LIHYASGENWIFCD	PMCKFIRFS	FFHNL	YSSIL
gi 6679193		VVISTYIFKMRPWKSSSTIIMLN	LACTDLYETSLP	LIHYASGENWIFCD	PMCKFIRFS	FFHNL	YSSIL
gi 4505557		VAIWMFVPHMKPWSGISVYMF	NLALADFLYVLT	LPALIFYYENKTDWIF	GDAMCKLQRF	FHVNL	YGSIL
gi 1352695		VAIWMFVPHMKPWSGISVYMF	NLALADFLYVLT	LPALIFYYENKTDWIF	GDAMCKLQRF	FHVNL	YGSIL
gi 1352691		VAIWMFVPHMKPWSGISVYMF	NLALADFLYVLT	LPALIFYYENKTDWIF	GDAMCKLQRF	FHVNL	YGSIL
		150	160	170	180	190	200	210
NOV3	
gi 16566323		FLTCISIFRYCVIITHMSCFS	SIHKTRCAVACAVVWII	SLVAVIEMTSLIT	STNRIN-RSAGLD	LTSSDE
gi 6679193		FLTCISIFRYCVIITHMSCFS	SIHKTRCAVACAVVWII	SLVAVIEMTSLIT	STNRIN-RSAGLD	LTSSDE
gi 4505557		FLTCISAHRYSGVVYPLKSL	GRLKKKNNAHYVSVL	VWLIWVVAISPILF	YSGTGIRKNKT	ITCYD	ITSDEY
gi 1352695		FLTCISAHRYSGVVYPLKSL	GRLKKKNNAHYVSVL	VWLIWVVAISPILF	YSGTGIRKNKT	ITCYD	ITSDEY
gi 1352691		FLTCISAHRYSGVVYPLKSL	GRLKKKNNAHYVSVL	VWLIWVVAISPILF	YSGTGIRKNKT	ITCYD	ITSDEY
		220	230	240	250	260	270	280
NOV3	
gi 16566323		LNTIKWYNLILATTFCEPL	VIVTLCTTIIHTL	THGLQDSC	LKORARRLIT	ILLALAFYVCF	PPFHRLR
gi 6679193		LNTIKWYNLILATTFCEPL	VIVTLCTTIIHTL	THGLQDSC	LKORARRLIT	ILLALAFYVCF	PPFHRLR
gi 4505557		LRSYFIYSMCTTVAMFC	PLVLILGCYGLIV	RALIKDLDNSPL	RRKSIYLVII	IVLTVFAVSYI	PPFHVMM
gi 1352695		LRSYFIYSMCTTVAMFC	PLVLILGCYGLIV	RALIKDLDNSPL	RRKSIYLVII	IVLTVFAVSYI	PPFHVMM
gi 1352691		LRSYFIYSMCTTVAMFC	PLVLILGCYGLIV	RALIKDLDNSPL	RRKSIYLVII	IVLTVFAVSYI	PPFHVMM
		290	300	310	320	330	340	350
NOV3	
gi 16566323		VIRTESRLLSIS	---CSIENQIHEAVI	VSRLAANTFGN	LLYVVVSDNEQ	QAVCSTVR	--CKVSGNLE
gi 6679193		VIRTESRLLSIS	---CSIENQIHEAVI	VSRLAANTFGN	LLYVVVSDNEQ	QAVCSTVR	--CKVSGNLE
gi 4505557		TMNLRARLDFQTFEMCA	ENDRVYATYQVTR	GLASLNSCVD	PILYFLAGDT	FRRRLSRATRK	KASRRSEANL
gi 1352695		TMNLRARLDFQTFEMCA	ENDRVYATYQVTR	GLASLNSCVD	PILYFLAGDT	FRRRLSRATRK	KASRRSEANL
gi 1352691		TMNLRARLDFQTFEMCA	ENDRVYATYQVTR	GLASLNSCVD	PILYFLAGDT	FRRRLSRATRK	KASRRSEANL
		360	370					
NOV3						
gi 16566323		QSKSEMTLNILSEFK	QNGDTS					
gi 6679193		QSKSEMTLNILSEFK	QNGDTS					
gi 4505557		QSKSEMTLNILSEFK	QNGDTS					
gi 1352695		QSKSEMTLNILSEFK	QNGDTS					
gi 1352691		QSKSEMTLNILSEFK	QNGDTS					

Table 3F lists the domain description from DOMAIN analysis results against NOV3. This indicates that the NOV3 sequence has properties similar to those of other proteins known to contain these domains.

Table 3F Domain Analysis of NOV3

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:62)
Length = 254 residues, 100.0% aligned
Score = 125 bits (315), Expect = 3e-30

```
NOV3: 50  GNAVVISITYIFKMRPWKSSSTIIMNLACTDLLYLTSPLFLIHYASGENWIFGDFMCKFI 109
      || +|| + + + + +||| |||+| +| || | +|+|| +|| +
00001: 1  GNLLVILVILRTKKLRTPNIFLLNLAVADLLFLLTLPWALYYLVGGDWVFGDALCKLV 60

NOV3: 110 RFSFHFNLYSILFLTCSIFRYCVIIHPMSCFSIHKTRCAVVACAVVWIIISLVAVIPMT 169
      | | +||| || || | +||+ | | | | +|+|+|+ +|
00001: 61  GALFVVNGYASILLTAISIDRYLAIVHPLRYRRIPTPRRAKVLILLVWVLALLSLPPL 120

NOV3: 170 FLITSTNRTNRSACDLTSSDELNTIKWYNLILTATTFCLPLVIVTLCYTTIIHTLTHGL 229
      + + | + + + | + | | |||+++ +||| | + ||
00001: 121 LFSWLRTEEGNTTVCLIDFPESVKRSYVLLSTLVGFVLPPLLVLVCYTRILRLRKRA 180

NOV3: 230 QTDSLCK-----QKARRLTILLLAFYVCFLPFHILRVIRIESRLSISCSIQIHEA 283
      ++ || +|| ++ +++++ | +|+|+|+ ++ || + |
00001: 181 RSQSLKRRSSSERKAAKMLLVVVVFLCWLPHYHIVLLLDL-----LCLLSIWRVLP 235

NOV3: 284 YIVSRPLAALNTFGNLLLY 302
      +++ || +| + ++|
00001: 236 LLITLWLAVVNSCLNPIIY 254
```

NOV3 is similar to the GPCR super family and in particular to the rhodopsin sub family and P2Y purinoceptor subtypes. G-protein-coupled receptors (GPCRs) constitute a vast protein family that encompasses a wide range of functions (including various autocrine, paracrine and endocrine processes). The rhodopsin-like GPCRs themselves represent a widespread protein family that includes hormone, neurotransmitter and light receptors, all of which transduce extracellular signals through interaction with guanine nucleotide-binding (G) proteins. Although their activating ligands vary widely in structure and character, the amino acid sequences of the receptors are very similar and are believed to adopt a common structural framework comprising 7 transmembrane (TM) helices. The cellular response to ATP are mediated by specific high-affinity receptors designated as P2 purinoceptors, five subclasses of which have been defined pharmacologically-P2X, P2Y, P2U, P2T, and P2Z. Because of the presence of the rhodopsin family GPCR domain and the homology to the purinoceptors, we anticipate that the novel sequence described here will have useful properties and functions similar to these genes. (Tokuyama et al., Cloning of rat and mouse P2Y purinoceptors. Biochem Biophys Res Commun 211(1):211-8, 1995; Leon et al., Cloning and sequencing of a human cDNA encoding endothelial P2Y1. Gene 171(2):295-7, 1996).

The above defined information for NOV3 suggests that this NOV3 protein may function as a member of a P2Y purinoceptor 1 protein family. Therefore, the NOV3 nucleic acids and proteins of the invention are useful in potential therapeutic and diagnostic applications. For example, a cDNA encoding the NOV3 protein may be useful in gene therapy, and the NOV3 protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from hyperparathyroidism, fertility, endometriosis, Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy and/or ARDS. The NOV3 nucleic acid encoding P2Y purinoceptor 1- like protein, and the P2Y purinoceptor 1- like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV4

NOV4 includes two novel LOMP-like proteins disclosed below. The disclosed proteins have been named NOV4a and NOV4b.

NOV4a

A disclosed NOV4a nucleic acid of 1508 nucleotides (designated CuraGen Acc. No. CG50255-01) encoding a novel LOMP-like protein is shown in Table 4A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 377-379 and ending with a TAA codon at nucleotides 1421-1423. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 4A, and the start and stop codons are in bold letters.

Table 4A. NOV4a Nucleotide Sequence (SEQ ID NO:7)

AGGAGAGAAGAAATGAAAAGCAGGCACTTGAGAAGTCTAAGAGAAGCTCTAAGACGTTTAAGGAAATGCTGCAGGACA
GGGAATCCCAAAATCAAAGTCTACAGTTCCGTCAGAAGGAGAATGTATTCTTTTGATGATGTGCTGGAGGAAGGAAA
GCGACCCCTTACAATGACTGTGTGAGAAGCAAGTTACCAGAGTGAGAGAGTAGAAGAGAAGGGAGCAACTTATCCTTCA
GAAATTCCTCAAGAGATTCTACCACTTTTGCAAAAAGAGAGGACCCGTGTAACTGAAATTCAGCTTCCTTCTCAA
AGTCCTGTGGAAGAACAAGCCAGCCTCTTTGCTTCTCTGCGTTTCACGGAGCACACAAATGGAATCAACTTGTGTTT
CAGCTTCTCTCCCGAAGTTACCGGAAAAGTACAGTCAGGTAAACATCTGTGGTCACACCAAGACCCTTGGCTC
TCAGACAAGGGGAATCTCATCACTCCCGAGATCTTACAGTATGGATGATGCTTGAAGTATAATGGAGATATTGAAGAC
ATTAAGAGAATCCAAACAATGTGGTCAGCACCCCTGCACCAAGCCCGGACGCAAGCCAACTGGCTTCAAGCTTATCTA
GCCAGAAAGAGGTAGCAGCAACAGAAGAAGATGTGACAAGGCTGCCCTCTCTACATCCCCCTTCTCATCTCTTCCCA
AGACCAGGCTGCCACTTCTAAAGCCACATGTCTTCCACATCTGGTCTTGATTTAATGTCTGAATCTGGAGAAGGGGAA
ATCTCCCAAGAGAAGTCTCAAGATCCAGGATCAGTTCAGTGATATGAGAATCAGCATAAACAGACGCTGGGA
AGAGTCTTGACTTTGGGTTTACAATAAAATGGGATATTCTGGGATCTTCGTAGCATCAGTTGAAGCAGGTAGCCAGC
AGAATTTTCTCAGTACAAGTAGATGATGAAATATTGCTATTAACAACACCAAGTTTTCATATAACGATTCAAAGAG
TGGGAGGAAGCCATGGCTAAGGCTCAAGAACTGGACACCTAGTGATGGATGTGAGGCGCTATGGAAGGCTGACTGG

GCAAAGACCAACCTTCCTGCCATTATACGGCATAAAACCTCAATCTCACCAGTATGGCTACCAAAATTATAGGTTCACTGAAACAAAGTGGATTGATGCAACTTCTGGAATTTACAACTCAGAAAAATCTTCAAATCTATCTGTAACAACTGATTTCTCCGAAAGCCTTCAGAGTTCTAATATGAATCCAAAGAAATCAATGGAATTCATGATGAAAGCAATGCITTTGAATCAAAAGCATCTGAATCCATTCTTTGAAAACTTAAAGGCGATCACAATTTTTGAACAAGGTAAACCACAAAGCTACAATTCATGCACCTTCATGGAATTGTTCTCCCTCTCCACTCTTCTCATGGTCTGTGGTGCTGGGCTACATTGTGTCTTCCCGTA

The nucleic acid sequence of NOV4a maps to chromosome 13 and has 1100 of 1100 bases (100%) identical to a *Homo sapiens* LOMP protein mRNA (gb:GENBANK-ID:AF144237|acc:AF144237.1) ($E = 4.3e^{-292}$).

- 5 A NOV4a polypeptide (SEQ ID NO:8) encoded by SEQ ID NO:7 is 348 amino acid residues and is presented using the one letter code in Table 4B. Signal P, Psort and/or Hydropathy results predict that NOV4a does not contain a signal peptide and is likely to be localized to the mitochondrial matrix space with a certainty of 0.5147 and the cytoplasm with a certainty of 0.4500. Cytoplasmic localization is more likely since PDZ domain-containing
- 10 proteins have been shown to participate in cellular junction formation, receptor or channel clustering, and intracellular signalling events (Ponting et al., PDZ domains: targeting signalling molecules to sub-membranous sites. Bioessays 1997 Jun;19(6):469-79).

Table 4B. NOV4a protein sequence (SEQ ID NO:8)

MESTCVSASLPRSYRKTDVRLTSVVTIPRPFQSQTRGISSLPRSYTMDDAWKYNIGDIEDIKRTPNNVVSTPAPSPDASQLASSLSSQKEVAATEEDVTRLPSPTSPFSSLSQDQAATSKATLSSTSGLDLMSSEGEISPOREVSRSQDQFSDMRISINQTPGKSLDFGFTIKWDIPGIFVASVEAGSPAEFSQLQVDDEIIAINTTKFSYNDSEKWEEMAKAQETGHLVMDVRRYKADWCKDQPSLPFIRHKTNLNLTSMATKIIGSPETKWIDATSGIYNSEKSSNLSVTTDFSESLQSSNIESKEINGIHDESNAPESKASESISLKNLKRRSQFFEQKQPOS

- 15 The NOV4a amino acid sequence has 262 of 329 amino acid residues (79%) identical to, and 278 of 329 amino acid residues (84%) similar to, a *Homo Sapiens* 797 amino acid residue LOMP protein (ptnr:SPTREMBL-ACC:Q9UQM5) ($E = 4.6e^{-126}$).

- NOV4a is expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney,
- 20 lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, Cochlea, Colon, Coronary Artery, Epidermis, Foreskin, Hair Follicles, Islets of Langerhans, Liver, Lung, Ovary, Thymus and Whole Organism. This information was derived by determining the tissue sources of the sequences that were included in the invention
- 25 including but not limited to SeqCalling sources, Public EST sources, Genomic Clone sources, Literature sources, and/or RACE sources.

In addition, NOV4a is predicted to be expressed in adult brain tissues because of the expression pattern of a closely related *Homo sapiens* LOMP protein mRNA homolog (gb:GENBANK-ID:AF144237|acc:AF144237.1).

NOV4b

5 A disclosed NOV4b nucleic acid of 1436 nucleotides (designated CuraGen Acc. No. CG50255-02) encoding a novel LOMP-like protein is shown in Table 4C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 21-23 and ending with a TAA codon at nucleotides 1374-1376. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 4C, and 10 the start and stop codons are in bold letters.

Table 4C. NOV4b Nucleotide Sequence (SEQ ID NO:9)

```
GCTCTAAGACGTTTAAGGAAATGCTGCAGGACAGGGAATCCCAAAATCAAAGTCTACAGTCCGTCAAGAAGGAGAAT
GTATTCITTTGATGATGTGCTGGAGGAAGGAAAGCGACCCCTACAATGACTGTGTGAGAGCAAGTTACAGAGTGAG
AGAGTAGAAGAGAAGGAGCAACTTATCCTTCAGAAATCCCAAGAAGATTCTACCACTTTTGCAAAAAGAGAGGACC
GTGTAACAACCTGAAATTCAGCTTCCTTCTCAAAGTCTGTGGAAGAACAAGCCAGCCTCTTTGTCTTCTGCGTTC
ACGGAGCACACAATGGAATCAACTCGTGTTCAGCTTCTCTCCCAAGATTACCGGAAAACGTATACAGTCAGGTTA
ACATCTGTGGTCAACCAAGACCCCTTTGGCTCTCAGACAAGGGAATCTCATCACTCCCAAGATCTTACAGATGGATG
ATGCTTGAAGTATAATGGAGATGTTGAAGACATTAAAGAGAACTCCAAACAATGTGGTCAGCACCCCTGCACCAAGCCC
GGACGCAAGCCAACCTGGCTTCAAGCTTATCTAGCCAGAAAGAGGTAGCAGCAACAGAAGAAGATGTGACAAGGCTGCC
TCTCCTACATCCCCCTTCTCATCTCTTCCCAAGACCGGCTGCCACTTCTAAAGCCACATTGTCTTCCACATCTGGTC
TTGATTTAATGCTCTGAATCTGGAGAAGGGGAAATCTCCCAACAAGAGAAGTCTCAAGATCCAGGATCAGTTTCACTGA
TATGAGAATCAGCATAAACCAGACGCTGGGAAGAGTCTTGACTTTGGGTTTACAATAAAATGGGATATTCTGGGATC
TTGCTAGCATCAGTTGAAGCAGGTAGCCAGCAGAAATTTCTCAGCTACAAGTAGATGATGAAATTATGCTATTAAACA
ACACCAAGTTTTCATATAACGATTCAAAGAGTGGGAGGAAGCCATGGCTAAGGCTCAAGAACTGGACACCTAGTGAT
GGATGTGAGGCGCTATGGAAGGCTGACTGGGGCAAGACCAACCTTCCCTGCCATTATACGGCATAAAACCCCTCAAT
CTCACCAGTATGGCTACCAAAATATAGGTTACCTGAAACAAAGTGGATGTGCAACTTCTGGAATTTCAACTCAG
AAAAATCTTCAAATCTGCTGTAAACAACCTGATTTCTCGAAAGCCTTCGGAGTCTTAATATTGAATCCAAAGAATCAA
TGAATTCATGATGAAGCAATGCTTTTGATTCAAAGCATCTGAATCCATTCTTTGAAAACTTAAAAAGGCGATCA
CAATTTTGAACAAGGTAAACCACAAGCTAAACAATTCATGCATTTTCATGGAATGTCTCCCTCTCCACTCTTCTCT
CATGGTCTGTGGTG
```

The nucleic acid sequence of NOV4b maps to chromosome 13 and has 1051 of 1054 bases (99%) identical to a *Homo sapiens* LOMP protein mRNA (gb:GENBANK-ID:AF144237|acc:AF144237.1) ($E = 5.4e^{-279}$).

15 A NOV4b polypeptide (SEQ ID NO:10) encoded by SEQ ID NO:9 is 451 amino acid residues and is presented using the one letter code in Table 4D. Signal P, Psort and/or Hydropathy results predict that NOV4b does not contain a signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.4500 and the mitochondrial matrix space with a certainty of 0.4475.

Table 4D. NOV4b protein sequence (SEQ ID NO:10)

```
MLQDRESQNKSTVPSRRRMYSFDDVLEEGKRPPTMTVSEASYQSERVEEKGATYPSEI PKEDSTTFKREDRVTEIQLPSQ
SPVEEQSPASLSSLSRSTOMESTRVSASLPRSYRKTDVRLTSVVTTPRPFQSOTRGISSLPSTMTDDAWKNGDVEDIKRT
PNNVVSTPAPSPDASQLASSLSQKEVAATEEDVTRLPSPTSPFSSLSQDQAATSKATLSSTSGLDLMSSESGEISPOREVS
RSQDQFSDMRISINQTPGKSLDFGFTIKWDIPGIFVASVEAGSPAEFSQLQVDDEI IAINNTKFSYNDSKWEEMAKAQETG
HLVMDVRRYKADWGDQPSLPPI RHKTLNLTSMATKI IGSPETKWIDATSGIYNSEKSSNLVTTDFSESLSRSSNIESKEIN
GIHDESNAFDSKASESISLKNLKRRSQFFEQGKQPS
```

The NOV4b amino acid sequence has 259 of 329 amino acid residues (78%) identical to, and 276 of 329 amino acid residues (83%) similar to, a *Homo Sapiens* 797 amino acid residue LOMP protein (ptnr:SPTREMBL-ACC:Q9UQM5) ($E = 3.9e^{-124}$).

NOV4b is expressed in at least brain tissues. Expression information was derived from the tissue sources of the sequences that were included in the derivation of NOV4b. In addition, NOV4b is predicted to be expressed in adult brain tissues because of the expression pattern of a closely related *Homo sapiens* LOMP protein mRNA homolog (gb:GENBANK-ID:AF144237|acc:AF144237.1).

Possible small nucleotide polymorphisms (SNPs) found for NOV4a are listed in Table 4E.

Table 4E: SNPs				
Variant	Nucleotide Position	Base Change	Amino Acid Position	Base Change
13376214	402	C > T	9	Ser > Phe

Possible small nucleotide polymorphisms (SNPs) found for NOV4b are listed in Table 4F.

Table 4F: SNPs			
Consensus Position	Depth	Base Change	PAF
208	19	C > T	0.105
1089	19	G > A	0.316

NOV4a and NOV4b are very closely homologous as is shown in the amino acid alignment in Table 4G.

Table 4G Amino Acid Alignment of NOV4a and NOV4b

	10	20	30	40	50	60	70
NOV4a
NOV4b	MLQDRESQNKSTVP	SRRRMYSFDDVLE	BGKRPTMTVSE	ASYQSERVEEK	GATYPSEI	PKEDSTT	FAKR
	80	90	100	110	120	130	140
NOV4a
NOV4b	EDRVITEIQLPSQ	SPVEEQSPASL	SSLRSRSTCM	ESTRVSASL	PRSYRKTD	TVRLTSV	VTPRPF
	150	160	170	180	190	200	210
NOV4a	IS	SLPRSYTMDD	AWKYN	GDIEDIK	RTPNNV	VSTPAP	SPDASQL
NOV4b	IS	SLPRSYTMDD	AWKYN	GDIEDIK	RTPNNV	VSTPAP	SPDASQL
	220	230	240	250	260	270	280
NOV4a	SS	LSQDQAAT	SKATLS	SSTSC	LDLMS	ESGE	GEIS
NOV4b	SS	LSQDQAAT	SKATLS	SSTSC	LDLMS	ESGE	GEIS
	290	300	310	320	330	340	350
NOV4a	PG	IFVAS	VEAGS	PAEFS	QLQVD	DEII	IAIN
NOV4b	PG	IFVAS	VEAGS	PAEFS	QLQVD	DEII	IAIN

NOV4b PGI FVASVEAGSPAEFSQLQVDDEIIAINNNTKFSYNDSKEWEEAMAKAQETGHLVMDVRRYKADWGKDQ

360 370 380 390 400 410 420

NOV4a PSLPFIRHKTLLNLTSMATKIIGSPETKWIDATSGIYNSEKSSNLSVTTDFSESLOSSNIESKEINGIHDE

NOV4b PSLPFIRHKTLLNLTSMATKIIGSPETKWIDATSGIYNSEKSSNLSVTTDFSESLSRSSNIESKEINGIHDE

430 440 450

NOV4a SNAFESKASESISLKNLKRRSQFFEQGKPOS

NOV4b SNAFESKASESISLKNLKRRSQFFEQGKPOS

Homologies to any of the above NOV4 proteins will be shared by the other NOV4 proteins insofar as they are homologous to each other as shown above. Any reference to NOV4 is assumed to refer to both of the NOV4 proteins in general, unless otherwise noted.

- 5 NOV4a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 4H.

Table 4H. BLAST results for NOV4					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7369019 ref NP_05349.2 (NM_005358)	LIM domain only 7 isoform a [Homo sapiens]	797	277/348 (79%)	277/348 (79%)	e-133
gi 14757689 ref XP_040875.1 (XM_040875)	LIM domain only 7 isoform a [Homo sapiens]	797	275/348 (79%)	276/348 (79%)	e-131
gi 7710131 ref NP_056667.1 (NM_015842)	LIM domain only 7 isoform b [Homo sapiens]	784	277/348 (79%)	277/348 (79%)	e-129
gi 7710133 ref NP_056668.1 (NM_015843)	LIM domain only 7 isoform c [Homo sapiens]	728	277/348 (79%)	277/348 (79%)	e-128
gi 14757685 ref XP_007068.4 (XM_007068)	hypothetical protein XP_007068 [Homo sapiens]	784	275/348 (79%)	276/348 (79%)	e-127

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 4I.

Table 4I ClustalW Analysis of NOV4a

- 1) NOV4a (SEQ ID NO:8)
- 2) gi|7369019|ref|NP_005349.2| (NM_005358) LIM domain only 7 isoform a [Homo sapiens] (SEQ ID NO:63)
- 3) gi|14757689|ref|XP_040875.1| (XM_040875) LIM domain only 7 isoform a [Homo sapiens] (SEQ ID NO:64)
- 4) gi|7710131|ref|NP_056667.1| (NM_015842) LIM domain only 7 isoform b [Homo sapiens] (SEQ ID NO:65)
- 5) gi|7710133|ref|NP_056668.1| (NM_015843) LIM domain only 7 isoform c [Homo sapiens] (SEQ ID NO:66)
- 6) gi|14757685|ref|XP_007068.4| (XM_007068) hypothetical protein XP_007068 [Homo sapiens] (SEQ ID NO:67)

10 20 30 40 50 60 70

NOV4a MESTCVSASLPRSYRKTDVRLTSVVTTPRPFGSQTRGISSLPRSYTMDDAWKYNQDIEDIKRTENNIVVST

gi|7369019| MESTCVSASLPRSYRKTDVRLTSVVTTPRPFGSQTRGISSLPRSYTMDDAWKYNQDIEDIKRTENNIVVST

gi|14757689| MESTRVASLPRSYRKTDVRLTSVVTTPRPFGSQTRGISSLPRSYTMDDAWKYNQDIEDIKRTENNIVVST

gi|7710131| MESTCVSASLPRSYRKTDVRLTSVVTTPRPFGSQTRGISSLPRSYTMDDAWKYNQDIEDIKRTENNIVVST

gi|7710133| MESTCVSASLPRSYRKTDVRLTSVVTTPRPFGSQTRGISSLPRSYTMDDAWKYNQDIEDIKRTENNIVVST

gi|14757685| MESTRVASLPRSYRKTDVRLTSVVTTPRPFGSQTRGISSLPRSYTMDDAWKYNQDIEDIKRTENNIVVST

640 650 660 670 680 690 700

NOV4a
gi|7369019| RSASVINKPEVSLPGIMRRGESLDNLSPRNSWROPPLNLNPTGFIYASSSVODFSRPPPOLVSTSNRAYM

```
gi |14757689| RSASVVKPEVPSLPGIMRRGESLDNLDSPRSNSWRQPPWLNQPTGFYASSSVQDFSRPPPQLVSTSNRAYM
gi |7710131| RSASVVKPEVPSLPGIMRRGESLDNLDSPRSNSWRQPPWLNQPTGFYASSSVQDFSRPPPQLVSTSNRAYM
gi |7710133| RSASVVKPEVPSLPGIMRRGESLDNLDSPRSNSWRQPPWLNQPTGFYASSSVQDFSRPPPQLVSTSNRAYM
gi |14757685| RSASVVKPEVPSLPGIMRRGESLDNLDSPRSNSWRQPPWLNQPTGFYASSSVQDFSRPPPQLVSTSNRAYM

      710           720           730           740           750           760           770
.....|.....|.....|.....|.....|.....|.....|.....|
NOV4a
-----
gi |7369019| RNPSSSVPPPSAGSVKTTSTTGVAATTQSPTPRSHSPSASQSGSQRLNRNSVSGKRICSYCNNIIGKGAAMI I
gi |14757689| RNPSSSVPPPSAGSVKTTSTTGVAATTQSPTPRSHSPSASQSGSQRLNRNSVSGKRICSYCNNIIGKGAAMI I
gi |7710131| RNPSSSVPPPSAGSVKTTSTTGVAATTQSPTPRSHSPSASQSGSQRLNRNSVPVSVTSEALPQEIKSSSETTN
gi |7710133| RNPSSSVPPPSAGSVKTTSTTGVAATTQSPTPRSHSPSASQSGSQRLNRN-----SWTANRH
gi |14757685| RNPSSSVPPPSAGSVKTTSTTGVAATTQSPTPRSHSPSASQSGSQRLNRNSVPVSVTSEALPQEIKSSSETTN

      780           790           800           810           820
.....|.....|.....|.....|.....|
NOV4a
-----
gi |7369019| ESLGLCYHLHCFKCVACDCLCGSSSGAEVRIRNHQCYCNDICYLRFKSGRPTAM
gi |14757689| ESLGLCYHLHCFKCVACDCLCGSSSGAEVRIRNHQCYCNDICYLRFKSGRPTAM
gi |7710131| CTA-----TTAISDSNLDCOPP-CEVSLHTKALLQIEEYVAAHVDL---
gi |7710133| VM-----
qi |14757685| CTA-----TTAISDSNLDCOPP-CEVSLHTKALLQIEEYVAAHVDL---
```

Table 4J lists the domain description from DOMAIN analysis results against NOV4a. This indicates that the NOV4a sequence has properties similar to those of other proteins known to contain these domains.

Table 4J Domain Analysis of NOV4a

```
gnl|Smart|smart00228, PDZ, Domain present in PSD-95, Dlg, and ZO-1/2.;
Also called DHR (Dlg homologous region) or GLGF (relatively well
conserved tetrapeptide in these domains). Some PDZs have been shown to
bind C-terminal polypeptides; others appear to bind internal (non-C-
terminal) polypeptides. Different PDZs possess different binding
specificities (SEQ ID NO:68)
Length = 86 residues, 82.6% aligned
Score = 48.9 bits (115), Expect = 5e-07
```

5

NOV4a: 169 FGFTIK---WDIPGIFVASVEAGSPAEFSQLQVDDEIIAINNTKFSYNDKSEWEEAMAK 224
 ||++ | + | + | || | + | + | | + + | | || +
 00228: 14 LGFSLVGGKDSGDGGVVSSVVPGPSAAKAGLKPGDVILEVNGTSVE--GLTHLEAVDL 70

NOV4a: 225 AQETG-HLVMDVRR 237
 + | | + + | |
 00228: 71 LKEAGGKVTLTVLR 84

LOMP is a protein that contains a single LIM domain and PDZ domain. It has been isolated from adult brain and its function is unknown. Given the large number of PDZ-containing proteins and wide range of possible binding specificities, it seems likely many transmembrane proteins, ion channels, and receptors will be organized and regulated by PDZ domain complexes. The complex anatomy of neurons demands a high degree of functional organization. Therefore, membrane receptors and ion channels are often localized to selected subcellular sites and coupled to specific signal transduction machineries. PDZ domains have come into focus as protein interaction modules that mediate the binding of a class of submembraneous proteins to membrane receptors and ion channels and thus subserve these organizational aspects.

PDZ (also called DHR or GLGF) domains are found in diverse membrane-associated proteins including members of the MAGUK family of guanylate kinase homologues, several protein phosphatases and kinases, neuronal nitric oxide synthase, and several dystrophin-associated proteins, collectively known as syntrophins. Many PDZ domain-containing proteins appear to be localised to highly specialised submembranous sites, suggesting their participation in cellular junction formation, receptor or channel clustering, and intracellular signalling events. PDZ domains of several MAGUKs interact with the C-terminal polypeptides of a subset of NMDA receptor subunits and/or with Shaker-type K⁺ channels. Other PDZ domains have been shown to bind similar ligands of other transmembrane receptors. The crystal structures of PDZ domains, with and without ligand, have been determined. These demonstrate the mode of ligand-binding and the structural bases for sequence conservation among diverse PDZ domains. Modular PDZ domains, found in many cell junction-associated proteins, mediate the clustering of membrane ion channels by binding to their C-terminus. The X-ray crystallographic structures of the third PDZ domain from the synaptic protein PSD-95 in complex with and in the absence of its peptide ligand have been determined at 1.8 angstroms and 2.3 angstroms resolution, respectively. The structures reveal that a four-residue C-terminal stretch (X-Thr/Ser-X-Val-COO(-)) engages the PDZ domain through antiparallel main chain interactions with a beta sheet of the domain. Recognition of the terminal carboxylate group of the peptide is conferred by a cradle of main chain amides provided by a Gly-Leu-Gly-Phe loop as well as by an arginine side chain. Specific side chain interactions and a prominent hydrophobic pocket explain the selective recognition of the C-terminal consensus sequence.

Several dozen signaling proteins are known to contain 80-100 residue PDZ repeats domains, several of which interact with the C-terminal tetrapeptide motifs X-Ser/Thr-X-Val-COO⁻ of ion channels and/or receptors. PDZ domains have been noted in mammals, flies, worms, yeast, plants, and bacteria. It has been suggested that two PDZ domains occur in bacterial high-temperature requirement A (htrA) and one in tail-specific protease (tsp) homologues, and that a yeast htrA homologue contains four PDZ domains. Sequence comparisons suggest that the spread of PDZ domains in these diverse organisms may have occurred via horizontal gene transfer. The known affinity of *Escherichia coli* tsp for C-terminal polypeptides is proposed to be mediated by its PDZ-like domain, in a similar manner to the binding of C-terminal polypeptides by animal PDZ domains. Experimental evidence using the genetics of *Drosophila*, *C. elegans*, and mice indicates that PDZ proteins are involved in the regulation of epithelial cell growth, differentiation, and morphogenetic

movements during development. These systems will undoubtedly continue to provide great insight into the role PDZ proteins play in these phenomena. However, the precise nature of the molecular complexes mediated by PDZ proteins in epithelial tissues is still unresolved, and this remains an area of active investigation.

5 The above defined information for NOV4 suggests that this NOV4 protein may function as a member of a LOMP protein family. Therefore, the NOV4 nucleic acids and proteins of the invention are useful in potential therapeutic and diagnostic applications. For example, a cDNA encoding the NOV4 protein may be useful in gene therapy, and the NOV4 protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, developmental and/or neurological disorders. Since experimental evidence using the genetics of *Drosophila*, *C. elegans*, and mice indicates that PDZ proteins are involved in the regulation of epithelial cell growth, differentiation, and morphogenetic movements during development, as well as in the interactions among the components of synaptic junctions (J Clin Invest, 103(6), 767-772, 1999; Neurosci Res 32(1):1-7, 1998). The NOV4 nucleic acid encoding LOMP-like protein, and the LOMP-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV5

20 A disclosed NOV5 nucleic acid of 1882 nucleotides (also referred to as 16467945_0_88_da1) encoding a novel Epidermal Growth Factor-like protein is shown in Table 5A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 243-245 and ending with a TAA codon at nucleotides 1851-1853. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 5A, and the start and stop codons are in bold letters.

Table 5A. NOV5 Nucleotide Sequence (SEQ ID NO:11)

GGACACTGACATGGACTGAAGGAGTAGAAAAGAGCGGGAGGGGGCTCCGGGCGCCGCGCAGCAGACCTACTCCGGCC
GCGCGCCTCGCCGCTGTCCTCCGGGAGCGGCAGCAGTAGCCCGGGCGGCGAGGGCTGGGGTTCTCTGAGACTCTCAGAGGG
GCGCCTCCCATCGGCGCCACCCACCCAACTGTTCTCTCGCGCCCACTGCGCTGCGCCCCAGGACCCGCTGCCCAACATGG
ATTTTCTCCTGGCGCTGGTGTCTGGTATCCTCGCTCTACCTGCAGGCGGCGCCGAGTTCGACGGGAGGTGGCCAGGCAAAAT
AGTGTGATCGATTGGCCTATGTCGTTATGGTGGGAGGATTGACTGCTGCTGGGGCTGGGCTCGCCAGTCTTGGGGACAGTGT
CAGCCTGTGTGCCAACCCAGATGCAAAATGGTGAATGTATCGGGCCAAACAAGTCAAGTGTGATCCTGGTTATGCTGGAA
AAACCTGTAATCAAGATCTAAATGAGTGTGGCCTGAAGCCCGGCCCTGTAAGCACAGGTGCATGAACACTTACGGCAGCTA
CAAGTGCTACTGTCTCAACGGATATATGCTCATGCTATATTGGAGGCAAATATCAATGTGATGACATAGACGAATGCTCACTT
CAGTATGGCTGTGATGTTGTTAAAGGACAAATACGGTGCCAGTGCCTATCCCTGGCCTGCAGCTGGCTCCTGATGGGAGGA
CCTGTGTAGATGTTGATGAATGTGCTACAGGAAGAGCCTCCTGCCCTAGATTAGGCAATGTGTCAACACTTTTGGGAGCTA
CATCTGCAAGTGTGATAAAGGCTTCGATCTCATGTATATTGGAGGCAAATATCAATGTGATGACATAGACGAATGCTCACTT
GTCAGTATCAGTGCAGCAGCTTGTCTCGATGTTATAACATACGTGGGTCTACAAGTCAAAATGTAAGAAGGATACCAAGG
GTGATGGACTGACTTGTGTATATCCCAAAAGTTATGATTGAACCTTCAGTCCAAATCATGTACCAAGGGAATGGTAC
CATTTTAAAGGGTGACACAGGAATAATAATTGGATTCTGATGTTGGAAGTACTTGGTGGCCTCCGAAGACACCATATATT
CCTCCTATCATTACCAACAGGCCTACTTCTAAGCCAACAACAGACCTACACCAAGCCAACCAATTCCTACTCCACCAC
CACCACCACCCCTGCCAACAGAGCTCAGAACACCTTACCACCTACAACCCAGAAAGGCCAACCCGAGTACAACATAT

AGCACCAGCTGCCAGTACACCTCCAGGAGGGATTACAGTTGACAACAGGGTACAGACAGACCCTCAGAAACCCAGAGGAGAT
GTGTTTCAGTGTCTGGTACACAGTTGTAATTTTGACCATGGACTTTGTGGATGGATCAGGGAGAAAGACAATGACTTGCCT
GGGAACCAATCAGGACCCAGCAGGTGGACAATATCTGACAGTGTGGCAGCCAAAGCCCCAGGGGAAAAGCTGCACGCTT
GGTGCTACCTCTCGGCCGCTCATGCATTACAGGGGACCTGTGCCTGTCTTACAGGCACAAGGTGACGGGGTGCCTCTGGC
ACACTCCAGGTGTTTGTGAGAAAAACAGGTGCCACGGAGCAGCCCTGTGGGGAAGAAATGGTGGCCATGGCTGGAGGCAAA
CACAGATCACCTTGGCAGGGGGCTGACATCAAGAGCGTCTCTTCAAAGGTGAAAAAGGCGTGGTCACACTGGGGAGATTGG
ATTAGATGATGTGAGCTTGAAGAAAGGCCACTGCTCTGAAGAACGCTAACTCCAGAACTAACAATGAACCTCTTAA

The NOV5 nucleic acid was identified on chromosome Xp22 and has 477 of 699 bases (68%) identical to a *Homo sapiens* epidermal growth factor repeat containing protein (EGFL6) mRNA (gb:GENBANK-ID:AF186084|acc:AF186084.1) ($E = 2.0e^{-54}$).

- 5 A disclosed NOV5 polypeptide (SEQ ID NO:12) encoded by SEQ ID NO:11 is 536 amino acid residues and is presented using the one-letter code in Table 5B. Signal P, Psort and/or Hydropathy results predict that NOV5 contains a signal peptide and is likely to be localized extracellularly with a certainty of 0.7475. The most likely cleavage site for a NOV5 peptide is between amino acids 19 and 20, at: AAA-EF.

Table 5B. Encoded NOV5 protein sequence (SEQ ID NO:12)

MDFLALVLVSSLYLQAAAEFDGRWPRQIVSSIGLCRYGGRIDCCWGWARSWGQCQPVCPCKKHGECIGPNKCKCHPGY
AGKTCNQDLNECGLKPRPCKHRCMNTYGSYKCYCLNGYMLMPDGGSCSSALTCSMANCQYGCDDVVKQIRCCQPSPLQLAP
DGRTCDVDDECATGRASCPRFRQCVNTFGSYICKCHKGFDMYIGGKYQCHDIDECSLGQYQCSSFARCYNIRGSYKCKCK
EGYQGDGLTCVYIPKVMIEPSGPIHVPKNGTILKGDGTGNNNWI PDVGSTWVWPCKTPYIPPIITNRPTSKPTTRPTPKPTP
IPTPPPPPLPTELRTPLPPTTPERPTTGLTTIAPAASTPPGGITVDNRVQTDPFQKPRGDVFSVLVHSCNFDHGLCGWIRE
KNDLHWEPIRDPAGGOYLTVSAKAPGGKAARLVPLPLRLMHSGLDCLSFHRKVTGLHSGTLQVVFVRKHGAHGAALWGRN
GGHGWROTQITLRGADI KSVVFKGEKRRGHTGEIGLDDVSLKKGHCSEER

10 The NOV5 amino acid sequence has 135 of 225 amino acid residues (60%) identical to, and 179 of 225 amino acid residues (79%) similar to, a *Homo Sapiens* 553 amino acid residue epidermal growth factor repeat containing protein (ptnr:SPTREMBL-ACC:Q9NZL7) ($E = 1.1e^{-112}$).

- 15 NOV5 is expressed in at least the following tissues: lung tumor, fetal lung, fetal skin, fetal umbilical cord, fetal liver/spleen and placenta. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, genomic clone sources, literature sources, and/or RACE sources.

- 20 Possible small nucleotide polymorphisms (SNPs) found for NOV5 are listed in Tables 5C and 5D.

Table 5C: SNPs

Consensus Position	Depth	Base Change	PAF
58	15	A > C	N/A
64	17	A > G	N/A
67	19	G > A	N/A
68	19	A > C	N/A
75	18	G > A	N/A

118	23	C > T	N/A
231	40	A > G	N/A
327	49	G > A	N/A
954	28	A > G	N/A

Table 5D: SNPs				
Variant	Nucleotide Position	Base Change	Amino Acid Position	Base Change
13376215	719	G > C	159	Gln > His
13376216	942	A > G	234	Ile > Val

NOV5 has homology to the amino acid sequences shown in the BLASTP data listed in Table 5E.

Table 5E. BLAST results for NOV5					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15430246 gb AAK96010.1 (AY035898)	nephronectin short isoform [Mus musculus]	561	245/273 (89%)	263/273 (95%)	e-147
gi 15430248 gb AAK96011.1 (AY035899)	nephronectin long isoform [Mus musculus]	578	245/290 (84%)	263/290 (90%)	e-144
gi 15128103 gb AAK84391.1 AF397007_1 (AF397007)	nephronectin [Mus musculus]	592	245/304 (80%)	263/304 (85%)	e-143
gi 15795193 ref NP_277060.1 (NM_033525)	nephronectin [Mus musculus]	609	245/321 (76%)	263/321 (81%)	e-140
gi 9506563 ref NP_062270.1 (NM_019397)	EGF-like-domain, multiple 6 [Mus musculus]	550	137/225 (60%)	180/225 (79%)	1e-85

5

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 5F.

Table 5F Clustal W Sequence Alignment

- 1) NOV5 (SEQ ID NO:12)
- 2) [gi|15430246|gb|AAK96010.1| \(AY035898\)](#) nephronectin short isoform [Mus musculus] (SEQ ID NO:69)
- 3) [gi|15430248|gb|AAK96011.1| \(AY035899\)](#) nephronectin long isoform [Mus musculus] (SEQ ID NO:70)
- 4) [gi|15128103|gb|AAK84391.1|AF397007_1 \(AF397007\)](#) nephronectin [Mus musculus] (SEQ ID NO:71)
- 5) [gi|15795193|ref|NP_277060.1| \(NM_033525\)](#) nephronectin [Mus musculus] (SEQ ID NO:72)
- 6) [gi|9506563|ref|NP_062270.1| \(NM_019397\)](#) EGF-like-domain, multiple 6 [Mus musculus] (SEQ ID NO:73)

	10	20	30	40	50	60	70
NOV5	MDLLALVIVSSLYLQAAAEFDGRWPR	---	QIVSSIGLCRYGGRIDCCGWARQSWGQCQF	---	---	---	---
gi 15430246	MAVLLAAVLASSLYLQVAADFGRWPR	---	QIVSSIGLCRYGGRIDCCGWARQSWGQCQF	---	---	---	---
gi 15430248	MAVLLAAVLASSLYLQVAADFGRWPR	---	QIVSSIGLCRYGGRIDCCGWARQSWGQCQF	---	---	---	---
gi 15128103	MAVLLAAVLASSLYLQVAADFGRWPR	---	QIVSSIGLCRYGGRIDCCGWARQSWGQCQF	---	---	---	---
gi 15795193	MAVLLAAVLASSLYLQVAADFGRWPR	---	QIVSSIGLCRYGGRIDCCGWARQSWGQCQF	---	---	---	---
gi 9506563	MQPPWGLALPLL	PWVTGGVGTSP	PDYGLSAIAHQPCVCCYGTMMACCGWKKNNKEVCEA	---	---	---	---
	80	90	100	110	120	130	140

[illegible]

gn1|Smart|smart00137, MAM, Domain in meprin, A5, receptor protein tyrosine phosphatase mu (and others); Likely to have an adhesive function. Mutations in the meprin MAM domain affect noncovalent associations within meprin oligomers. In receptor tyrosine phosphatase mu-like molecules the MAM domain is important for homophilic cell-cell interactions. (SEQ ID NO:74)

Length = 163 residues, 96.9% aligned
Score = 86.3 bits (212), Expect = 4e-18

NOV5:	391	HSCNFDHG-LCGWIREKNDNLHWE-----PIRDPAGGQ--YLTVSAAKAPGGKAA	437
		+ + + ++ ++ + + + +	
00137:	4	GNCDFEEGNTCGWHQDSNDGPPERVSSATRNDGPNRDHTTGNHGHYMFETSSGKPGQTA	63
NOV5:	438	RLVLPLGLRLMHSGDLCLSFHVKVTLGHSGLQVFRKHGA-HGAALWGRNG--GHGWRQT	494
		+ + + + + + +	
00137:	64	RLSPPLYENRSTH-CLTFWYYMYGSGVGTNLVYVRVNGPQDTLLWSRSGTQGGQWLQA	122
NOV5:	495	QITLRGADIK-SVVFKEKRRGHTGEIGLDDVSLKKGHC	532
		++ + + + +	
00137:	123	EVALSTSPOPFOVVFEGTRGGGPGSYIALDDILLSNGPC	161

grrl Pfam|pfam00629, MAM, MAM domain.. An extracellular domain found in many receptors. (SEQ ID NO:75)
Length = 159 residues, 100.0% aligned
Score = 84.3 bits (207), Expect = 2e-17

NOV5:	393	CNFDHGL-CGWIREKDNDLHWE-----PIRDPAAGQ--YLTVSAAKAPGGKAAR	438
		+ + ++ + + + +	
00629:	1	CDFEDGSHCGWSQDSGDDLWTRVNSATGGSTGPRGDHTTGNGHYMYVDTSSEGQGTAR	60
NOV5:	439	LVLPLGRLMHSGDLCLSFRHKVTGLHSGTLQVVRKHGAHGA-ALWGRNG--GHGWRQTQ	495
		+ + + + + ++ + +	
00629:	61	LLSPPLPPKRSP-CCLTFWYHMYGSGVGTLNVVYVRENGGPSDRLLWSRSGHGQGSWLLAE	119
NOV5:	496	ITLRGADIKS-VVFKGEKRRGHTGEIGLDDVSLKKGHCSE	534
		+ + + + + + + ++	
00629:	120	VTLPTSTKPFQVFEETRGGSGRGGIALDLSLSEGPCNO	159

gsl|Smart|smart00179, EGF_CA, Calcium-binding EGF-like (SEQ ID NO:76)
Length = 41 residues, 100.0% aligned
Score = 41.2 bits (95), Expect = 2e-04

NOV5: 214 DIDECSLGQYQCSSFARCYNIRGSYKCK-CKEGYQGDGLTCV 254
 |||||+| | +. | | |||+|+ | | | |
 00179: 1 DIDECASGNP-CQGGTCVNTVGSYRCEECPGYTLDRNCE 41

41

EGF is a single-chain polypeptide consisting of 53 amino acids and having a molecular mass of about 6,000. Urdea et al. (Proc. Nat. Acad. Sci. 80: 7461-7465, 1983) synthesized the gene for human EGF.

By the study of human-rodent somatic cell hybrids with a genomic DNA probe,
5 Brissenden et al. (Am. J. Hum. Genet. 36: 133S only, 1984) mapped the EGF locus to 4q21-4qter, possibly near TCGF, the locus coding for T-cell growth factor (147680). Both nerve growth factor (see NGFB, 162030) and EGF are on mouse chromosome 3 but they are on different chromosomes in man: 1p and 4, respectively (Zabel et al., Proc. Nat. Acad. Sci. 82: 469-473, 1985). Zabel et al. (1985) pointed out that mouse chromosome 3 has one segment
10 with rather extensive homology to distal 1p of man and a second with homology to proximal 1p of man. By in situ hybridization, Morton et al. (Cytogenet. Cell Genet. 41: 245-249, 1986) assigned EGF to 4q25-q27. The receptor for EGF (EGFR; 131550) is on chromosome 7.

The EGF repeat superfamily of genes often encodes proteins that govern cellular proliferative responses (Yeung G, et.al.; Genomics 1999 Dec 1;62(2):304-7). Using a high-
15 throughput screening by hybridization approach, a novel human EGF repeat superfamily member that maps to human chromosome X was identified. Termed EGFL6, the gene encodes a predicted signal peptide, suggesting that it is secreted. Other predicted features include four and one-half EGF-like repeat domains, two N-linked glycosylation sites, an integrin association motif (RGD), and a tyrosine phosphorylation site. Importantly, its transcripts are
20 expressed in brain and lung tumor and fetal tissues, but are generally absent from normal adult tissues. Implications with respect to cell cycle regulation and oncogenesis have been suggested.

EGF repeat motif defines a superfamily of diverse proteins involved in regulating a variety of cellular and physiologic processes. This motif features a series of conserved
25 cysteines and glycines positioned in a domain of 30 to 40 residues. EGF-like repeat family members are predominantly secreted or cell surface molecules, often involved in the regulation of cell cycle, proliferation, and developmental processes. Using a high-throughput screening-by-hybridization approach, Yeung et al. (1999) identified the EGFL6 gene. The predicted 553-amino acid EGFL6 protein has a putative N-terminal signal peptide, which
30 suggests that it is secreted; an EGF repeat region containing 4 complete EGF-like repeats and 1 partial EGF-like repeat; an integrin association motif (RGD); 2 potential N-glycosylation sites; and a potential tyrosine phosphorylation site. Northern blot analysis of a variety of normal human tissues detected an approximately 2.4-kb EGFL6 transcript only in placenta. Among the cancer tissues tested, EGFL6 expression was found only in meningioma tumors.

Screening-by-hybridization analysis of various cDNA libraries indicated EGFL6 expression in lung tumor, fetal lung, fetal skin, fetal umbilical cord, fetal liver/spleen, and placenta, but not in normal adult tissues, including lung. By analysis of a somatic cell hybrid mapping panel, Yeung et al. (1999) mapped the EGFL6 gene to chromosome X. They noted that a UniGene cluster corresponding to the EGFL6 gene contains an STS that has been mapped to Xp22. Epidermal growth factor (EGF) repeat-containing proteins constitute an expanding family of proteins involved in several cellular activities such as blood coagulation, fibrinolysis, cell adhesion, and neural and vertebrate development (Buchner G, et.al.; Genomics 2000 Apr 1;65(1):16-23).

EGF is produced in abundance by the mouse submandibular gland. Tsutsumi et al. (Science 233: 975-977,1986) found that sialoadenectomy decreased circulating EGF to levels below detection but did not affect testosterone or FSH levels. At the same time a decrease in spermatids in the testis and mature sperm in the epididymis decreased. The changes were corrected by administration of EGF. A role of EGF in some cases of human male infertility, particularly those with unexplained oligospermia, was proposed.

During the immediate-early response of mammalian cells to mitogens, histone H3 (see 601128) is rapidly and transiently phosphorylated by one or more kinases. Sassone-Corsi et al. (Science 285: 886-891, 1999) demonstrated that EGF-stimulated phosphorylation of H3 requires RSK2 (300075), a member of the pp90(RSK) family of kinases implicated in growth control.

EGF repeat-containing proteins constitute an expanding family of proteins involved in several cellular activities such as blood coagulation, fibrinolysis, cell adhesion, and neural and vertebrate development. By using a bioinformatic approach, Bucher et al. have identified a new member of this family named MAEG (MAM- and EGF-containing gene; HGMW-approved gene symbol and gene name). Sequence analysis indicates that MAEG encodes a secreted protein characterized by the presence of five EGF repeats, three of which display a Ca(2+)-binding consensus sequence. In addition, a MAM domain is also present at the C-terminus of the predicted protein product. The human and murine full-length cDNAs were identified and mapped to human Xp22 and to the mouse syntenic region. Northern analysis indicates that MAEG is expressed early during development. Taken together, these data render MAEG a candidate for human and murine developmental disorders. (Buchner G, et.al.; Genomics 2000 Apr 1;65(1):16-23).

The above defined information for NOV5 suggests that this NOV5 protein may function as a member of a Epidermal Growth Factor-like protein family. Therefore, the NOV5

5 Craniofrontonasal dysplasia; Deafness, X-linked 6, sensorineural; Goiter, multinodular, 2;
Mental retardation, X-linked nonspecific, 58; Opitz G syndrome, type I; Partington syndrome
II; Simpson-Golabi-Behmel syndrome, type 2; Simpson-Golabi-Behmel syndrome, type 2;
Oncogenesis; fertility; regulation of cell cycle, proliferation and developmental processes.
The NOV5 nucleic acid encoding the Epidermal Growth Factor-like protein, and the
10 Epidermal Growth Factor-like protein of the invention, or fragments thereof, may further be
useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the
protein are to be assessed.

NOV6 includes three novel Hyaluronan-mediated Motility Receptor-like proteins
15 disclosed below. The disclosed proteins have been named NOV6a, NOV6b and NOV6c.

A disclosed NOV6a nucleic acid of 2684 nucleotides (also referred to as CG50239-01) encoding a novel Hyaluronan-mediated Motility Receptor-like protein is shown in Table 6A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 37-39 and ending with a TAA codon at nucleotides 2164-2166. Putative untranslated regions upstream from the start codon and downstream from the termination codon are underlined in Table 6A, and the start and stop codons are in bold letters.

GCCAGTCACTTTCAGTTTCTGGCGCTGGCCGTCACAACTGTCTTTCTTAAGGCGCCCTTGAAACGATTCAATGACCCTTCTG
 GTTGTGCAACCTTCCAGGTGCTCTATGATGTTAAATCTTTAGAAGATTGAAGAGCCAGTATCTCTTCAGAAATCAACAAG
 ATTTAAACCAAAAAAGCTCAATCAAAACAACTCTAATGTTGCAAAAGATACCTACCTTGCTGCTTCAGTCAAGAAAGTTAAG
 TCTTCGGAATCAAAGATTCTGTGTTCTTCTACAGGAACGTGGTGCCAGGACAGGCGGATCCAGGATCTGGAAACTGAGTTGG
 AAAAGATGGAAGCAAGGCTAAATGTCTGCATCAAGGGAAGAAAACTCTCTCTGCAAAATATGTCTACCTGGAAAAACAACCT
 TATTGAATTGACCAAGCTAATGAACATACTAAATCTAAGTTTCTGAAATGTTAACCAGAAGAA'TTGAAGATTTCAAGC
 TTGGAGTTGATGAAACTTAGAAACAAAAGAGAAAACAAAGATGAGGGGTATGATGGCTAAGCAAGAAGGCATGGAGATGAAGC
 TGCAGGTCACCCAAAAGAGGCTCGAAGAGTCTCAAGAGAAATACCCAACTGGAGGGAAAACTGTGTTCAATGAGAAAAAGA
 AAGATTGATGAAAAATCTGAAACCAAGAAAACTCTTGGAAATACCTGAAGAATTTAGTTGTGCTTCAGATCAAGTGGAAAAA
 TACAAGCTAGATTATGCCCAGTTAGAAGAAAAATTGAAAGAGAAGATGATGAAATTTTAAGCCTTAAGCAGTCTCTTGAGG
 ACAATATTGTATATTATCTAAACCAAGTAGAAGTCAAAATGTAAGTTGCAAGTCTGAAACCAAGAAAAAGACCACTGT
 CAACAGGAATGAGAAACAACAACGAAATCAATGCAGAGATCGAAACTAGACGAGGTTATTCTTGAAACACGGGA
 CATGAAAAGCTTCAACAAAAAGAAATTACAAATTTGATTCACTTCTGCAACAAGAGAAAGAAATTACTTCGAGTCTTCATCAGA
 AGTCTGTGTTCTTTCAAGAGGAAATGGTTAAAGAGGAAGATCTGTTGAGGAGAATTAAGAGCAAACTCGATGAGCTGTA
 TAAATTACAGCAAAAGGAGGAACAAGCTGAAAGGCTGGTCAAGCAATTTGAAGAGGAAGCAAAATCTAGACTGAAAGATTA
 AAACCTCTAGAAGAAAAGCTGAAAGGGAAGGAGGCTGAAGCTGGAGAAAAGTAGTGCTGCTCATACCCAGGCCACCTCGCTTT
 TGCAGAAAAGATGATGACGATGTGGTCAAGGCTGAAGATGTTACTGCTCAATTTGAAGAGCTATAAGCGTTAAACGCCAG
 TGAGATGAGAATCTTAAGCTGGAGACTCATCTTACAGAAAAGCGGCCAAGCTGGGAAAAATGCAAGAGATGTTGAC
 CATCAGATTTTGGCAACTGAGAGCTCAAAATCAAGAATATGTAAGGATGCTTCTAGATCTGCAGACCAAGTCAGCACTAAAGG
 AAACAGAAATTAAGAAATACAGTTTCTTTCTCAAAAATTAAGTATTGTCGAAACCACTCAAGCAACAGGAGGAAGA
 CTTTAGAAAACTGAGTGAAGATGAAGGAAGCAAGAAAGCTGAAAAGAGAAATTAACACAGCAGAATTAAGTGAAGAAATTAAC

AAGTGGCGTCTCCTCTATGAAGAACTATATAATAAAACAAAACCTTTTCAGCTACAACCTAGATGCTTTTGAAGTAGAAAAAC
AGGCATTGTTGAATGAACATGGTGCAGCTCAGGAACAGCTAAATAAAATAAGAGATTTCATATGCTAAATTATTTGGGTTCATCA
GAATTTGAAACAAAAAATCAAGCATGTTGTGAAGTTGAAAGATGAAAAATAGCCAACCTCAAATCGGAAGTATCAAACCTCCGC
TGTCAGCTTGCTAAAAAACAAGTGAGACAAAACCTCAAGAGGAATTGAATAAAGTTCTAGGTATCAAACACTTTGATC
CTTCAAAGGCTTTTCATCATGAAAGTAAAGAAAATTTGCCCCGAAGACCCCATTAAGAAGGCAATACAAACTGTTACCG
AGCTCCTATGGAGTGTCAAGAATCATGGAAGTAAACATCTGAGAAACCTGTTGAAGATTATTTTCATTGCTCTGTGTATT
GATGTTGCTGTTATTATATTGACATGGGTATTTTATAATGTTGTATTTAATTTTAACTGCCAATCCTTAAATATGTGAAG
GAACATTTTACCAAAGTGCTTTTACATTTTATTTTCTTGCAAATACCTCCTCCCTAATGCTCACCTTTATCACCTC
ATTCTGAACCCCTTTGCTGGCTTTCCAGCTTAGAATGCATCTCATCAACTTAAAGTCAGTATCATATTATATCCTCCTGT
TCTGAAACCTTAGTTTCAAGAGTCTAAACCCAGATTCTTCAGCTTGATCCCTGGAGGCTTTTCTAGTCTGAGCTCTTTAGC
TAGGCTAAACACCTTGGCTTGTATTGCTCTACTTTGATTCTTGATAATGCTCACTTGGTCTACCTATTATCCTTTCTA
CTTGTCAGTTTCAAAATAAGAAATAAGGACAAGCCTAATTCATAGTAACCTCTCTATTTT

The NOV6a nucleic acid was identified on chromosome 5 and has 2444 of 2453 bases (99%) identical to a *Homo sapiens* hyaluronan receptor (RHAMM) mRNA (gb:GENBANK-ID:HSU29343|acc:U29343.1) (E = 0.0).

- 5 A disclosed NOV6a polypeptide (SEQ ID NO:14) encoded by SEQ ID NO:13 is 709 amino acid residues and is presented using the one-letter code in Table 6B. Signal P, Psort and/or Hydropathy results predict that NOV6a does not contain a signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500.

Table 6B. Encoded NOV6a protein sequence (SEQ ID NO:14)

MSFPPKAPLKRFPNDPSGCAPSPGAYDVKTLEVLKGPVSFQKSRFKQKESKQNLNVKDTTLPASARKVKSSESIRVLLQ
ERGAQDRRIQDLETELEKMEARLNAALREKTSLSANNATLEKQLIELTRTNELLKSKFSENGNQKNLRILSLELMKLRNKR
ETKMRGMMAKQEGMEMKLQVTQSRLEESQGGKIAQLEGKLVSEKEKIDKSETEKLLLEYIEEISCASDQVEKYKLDIAQLE
ENLKEKNDIELSLKQSLLEDNIVILSKQVEDLNVKQQLLETEKEDHVNRRNREHNENLNAEMQNLEQKFI LEQREHEKLQKE
LQIDSLLOQEKELSSSLHQKLCSPQEEVMVEKKNLFEEELKQTLDELDKLQKEEQAEERLVKQLEEEAKSRAEELKLEEKI
KGKEAELEKSSAHTQATLLQEKYDSMVQSLQEDVTAQFESYKALTASEIEDLKLENSSLQEKAAKAGKNAEDVQHILAT
ESSNQEYVRMLLDLQTKSALKETEIKEITVSPLOKITDLQNLKQEEEDFRKQLEDEEGRKA EKENTTAELTEEINKWRL
YEELYNKTKPFQLQDAFEVEKQALLNEHGAQEQNLKIRDSYAKLLGHQNLKQKI KHVVKLKDENSQKSEVSKLRCLQA
KKKQSETKLQEBLNKVLGIKHFDPSKAFHHESKENFALKTPLKEGNTNCYRAPMECQESWK

- 10 The NOV6a amino acid sequence has 703 of 724 amino acid residues (97%) identical to, and 706 of 724 amino acid residues (97%) similar to, a *Homo sapiens* 724 amino acid residue Hyaluronan mediated motility receptor (intracellular Hyaluronic acid binding protein) (receptor for Hyaluronan-mediated motility) protein (ptnr:SWISSPROT-ACC:O75330) (E = 0.0).

- 15 NOV6a is expressed in at least the following tissues: bone marrow, brain, colon, coronary artery, epidermis, liver, lung, lymph node, mammary gland/breast, ovary, placenta, prostate, stomach, testis, tonsils, uterus and whole organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE
20 sources.

NOV6b

A disclosed NOV6b nucleic acid of 2020 nucleotides (also referred to as CG50239-02) encoding a novel Hyaluronan-Mediated Motility Receptor-like protein is shown in Table 6C.

An open reading frame was identified beginning with an ATG initiation codon at nucleotides 36-38 and ending with a TAA codon at nucleotides 1974-1976. Putative untranslated regions upstream from the start codon and downstream from the termination codon are underlined in Table 6C, and the start and stop codons are in bold letters.

Table 6C. NOV6b Nucleotide Sequence (SEQ ID NO:15)

GCAGTCACCTTCAGTTTCTGGAGCTGGCCGTCAACATGTCCTTTCCTAAGGCGCCCTTGAAACGATTCAATGACCCCTTC
 TGGTTGTGCACCATCTCCAGGTGCTTATGATGTTAAACTTTAGAAAGTATTGAAAGGACCAAGTATCCTTTCAGAAATCA
 CAAAGATTTAAACAACAAAAGAATCTAAACAAAATCTTAATGTTGACAAAGATACTACCTTGCGCTGCTTCAGCTAGAA
 AAGTTAAGTCTTCGGAATCAAAGATTGTGTTCTTCTACAGGAACGTGGTGCCAGGACAGGCGGATCCAGGATCTGGA
 AACTGAGTTGGAAAAGATGGAAGCAAGGCTAAATGCTGCACTAAGGGAAAAACATCTCTCTGCAAAATAATGCTACA
 CTGGAAAAACAACCTTATGAAATTGACCAGGACTAATGAACTACTAAAACTAAGGTTTCAATAGAGAAAAAGAAAGATTG
 ATGAAAAATCTGAAACAGAAAAACTCTTGGAAATACATCGAAGAAATTAGTTGTGCTTCAGATCAAGTGGAAAAATACAA
 GCTAGATATTGCCAGTTAGAAGAAAATTGAAAGAGAAGAAATGATGAAATTTAAGCCTTAAGCAGTCTCTTGAGGAC
 AATATTGTTATATCTAAACAAGTAGAAGATCTAAATGTGAAATGTGAGTGTGAAACAGAAAAAGAACCATG
 TCAACAGGAATAGAGAACACAAAGAAATCTAAATGCAGAGATGCAAACTTAGAACAGAAAGTTATTCTTGAACAACG
 GGAACATGAAAAGCTTCAACAAAAAGAATTACAAATTGATTCACTTCTGCAACAAGAGAAAGAAATTATCTTCGAGTCTT
 CATCAGAAGCTCTGTTCTTTCAAGAGGAAATGGTTAAAGAGAAGAAATCTGTTTGAGGAAGAAATTAAAGCAAACTGG
 ATGAGCTTGATAAATTACAGCAAAAGGAGGAACAAGCTGAAAGGCTGGTCAAGCAATTGGAAGAGGAAGCAAAATCTAG
 AGCTGAAGAATTAAGCTCTAGAGAAAAGCTGAAAGGGAAGGAGGCTGAACTGGAGAAAAGTAGTGCTGCTCATACC
 CAGGCCACCCCTGCTTTTGCAGGAAAAGTATGACAGTATGGTGCAAGCCTTGAAGATGTTACTGCTCAATTTGAAAGCT
 ATAAAGCGTTAACAGCCAGTGAGATAGAAGATCTTAAGCTGGAGAACTCATCATTACAGGAAAAAGCGGCCAAGGCTGG
 GAAAAATGCAGAGGATGTTGAGCATCAGATTTTGGCAACTGAGAGCTCAAATCAAGAAATATGTAAGGATGCTTCATAG
 CTGCAGACCAAGTCAGCACTAAAGGAACAGAAATTAAGAAATCACAGTTTCTTTCTTCAAAAAATAACTGATTGTC
 AGAACCAACTCAAGCAACAGGAGGAGAGCTTTAGAAAACAGCTGGAAGATGAAGAAGGAAGAAAAGCTGAAAAAGAAAA
 TACAACAGCAGAATTAAGTGAAGAAATTAACAAGTGGCGTCTCCTCTATGAAGAACTATATAATAAAACAAACCTTTT
 CAGCTACAACATAGATGCTTTTGAAGTAGAAAAACAGGCATTGTTGAATGAACATGGTGCAGCTCAGGAACAGCTAAATA
 AAATAAGAGATTCATATGCTAAATATTGGGTGTCAGAAATTTGAAACAAAAAATCAAGCATGTTGTGAAGTTGAAAGA
 TGAAAAATAGCCAACTCAAATCGGAAGTATCAAACTCCGCTGTGAGCTTGTCTAAAAAACAAGTGAAGCAAACTT
 CAAGAGGAATTGAATAAAGTTCTAGGTATCAACACTTTGATCCTTCAAGGCTTTTCATCATGAAAGTAAAGAAAATT
 TTGCCCTGAAGACCCCATTAAGAAAGGCAATACAACTGTTACCGAGCTCTATGGAGTGTCAAGAAATCATGGAAGTA
 AACATCTGAGAAAACCTGTTGAAGATTATTTCATTCGTTCTGTTGT

The NOV6b nucleic acid was identified on chromosome 5q32 and has 1571 of 1571 bases (100%) identical to a *Homo sapiens* hyaluronan receptor (RHAMM) mRNA (gb:GENBANK-ID:HSU29343|acc:U29343.1) (E = 0.0)

A disclosed NOV6b polypeptide (SEQ ID NO:16) encoded by SEQ ID NO:15 is 646 amino acid residues and is presented using the one-letter code in Table 6D. Signal P, Psort and/or Hydropathy results predict that NOV6b does not contain a signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500.

Table 6D. Encoded NOV6b protein sequence (SEQ ID NO:16).

MSFPKAPLKRFPNDPSGAPSPGAYDVKTLEVLKGPVSFQKSQRFKQKESKQNLNVDKDTTLPASARKVKSSSESKI CVLLQERGA
 QDRRIQDLETELEKMEARLNALREKTSLSANNATLEKQLIELTRTNELLKSKVSI EKEKIDEKSETLEKLEYIEEISCASDQVE
 KYKLDIAQLEENLKEKNDEILSLKQSLLEDNI VILSKQVEDLNVKQQLLETEKEDHVNRRNREHNENLNAEMQNLEQKFI LEQREHE
 KLQKELQIDSLQKEKELSSSLHQKLCSFQEEVMKEKNLFEELKQTLDELKQKBEQAERLVKQLEEEAKSRAEELKLEE
 KLKGKEAELEKSSAAHTQATLLQEKYDSMVQSLEDVTAQFESYKALTASEIEDLKLENSSLQEKAAKAGKNAEDVQHQILATES
 SNQEVYRMLLDLQTKSALKETEIKEITVSFLQKITDLQNLKQOEEDFRKQLEDEEGRKAKEKNTTAELETIEINKWRLLYEELYN
 KTKPFQLQDAFEVEKQALLNEHGAAQEQNLNKRDSYAKLLGHQNLKQKIKHVVKLKDENSQKSEVSKLRCLAKKKQSETKLQ
 EELNKVLGIKHFDPSKAFHHESKENFALKTPLKEGNTNCRYAPMECQESWK

The NOV6b amino acid sequence has 527 of 626 amino acid residues (84%) identical to, and 555 of 626 amino acid residues (88%) similar to, a *Homo sapiens* 725 amino acid residue hyaluronan receptor protein (ptnr:pir-id:JC5016) (E = 7.6e⁻²⁶³).

NOV6b is expressed in at least the following tissues: bone marrow, brain, colon,

coronary artery, epidermis, liver, lung, lymph node, mammary gland/breast, ovary, placenta, prostate, stomach, testis, tonsils and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

5 NOV6c

A disclosed NOV6c nucleic acid of 2187 nucleotides (also referred to as CG50239-03) encoding a novel Hyaluronan-Mediated Motility Receptor-like protein is shown in Table 6E. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 37-39 and ending with a TAA codon at nucleotides 2164-2166. Putative untranslated regions upstream from the start codon and downstream from the termination codon are underlined in Table 6E, and the start and stop codons are in bold letters.

Table 6E. NOV6c Nucleotide Sequence (SEQ ID NO:17)

```
GCCAGTCACCTTCAGTTTCTGGAGCTGGCCGTCAACATGTCCTTTCTAAGGCGCCCTTGAAACGATTCAATGACCCTT
CTGGTTGTGCACCATCTCCAGGTGCTTATGATGTTAAACCTTAGAAGTATTGAAAGGACCATCTCTTCAGAAATC
ACAAAGATTTAAACAACAAAAAGAATCTAAACAAATCTTAATGTTGACAAAGATACTACCTTGCTGCTTCAGCTAGA
AAAGTTAAGTCTTCGGAATCAAAGATTCTGTTCTTCTACAGGAACGTGGTGCCAGGACAGGCGGATCCAGGATCTGG
AAACTGAGTTGGAAGATGGAAGCAAGGCTAAATGCTGCACTAAGGGAAAAAACATCTCTCTGCAATAATGCTAC
ACTGGAACAACTTATGAAATGACCAAGGACTAATGAACTACTAAAACTAAGTTTCTGAAATGATAACCAAG
AATTTGAGAATTTAAGCTTGGAGTTGATGAACTTAGAAACAAAGAGAAACAAAGATGAGGGGTATGATGGCTAAGC
AAGAAGGCATGGAGATGAAGCTGCAGGTCAACCAAGGAGTCTCAAGAGTCTCAAGGAAAAATAGCCCACTGGAGGG
AAAACTTGTTCATAGAGAAAGAAAGATTGATGAAAAATCTGAAACAGAAAACTCTTGAATACATCGAAGAAAT
AGTTGTGCTTCAGATCAAGTGGAAAAATACAAGCTAGATATTGCCAGTTAGAAGAAATTTGAAAGAGAAGATGATG
AAATTTTAAGCCTTAAGCAGTCTCTTGAGGAGAAATATTGTTATATTATCTAAACAAGTAGAAGATCTAAATGTGAAATG
TCAGCTGCTTGAAAAAGAAAAAGAACCATGTCAACAGGAATAGAGAACACACGAAAACTAAATGCAGAGATGCAA
AACTTAAACAGAAGTTTATTCTTGAACACAGGAACGTGAAAGCTTCAACAAAAAGAAATACAAATTGATTCACCTTC
TGCAACAGAGAAAGAAATATCTTCGAGTCTTCATCAGAAGCTCTGTTCTTTTCAAGAGGAAATGGTTAAAGAGAAGAA
TCTGTTTGAGGAAGAATTAAAGCAAACTGGATGAGCTTGATAAATTACAGCAAAAGGAGGAACAAGCTGAAAGGCTG
GTCAAGCAATTGGAAGAGGAAGCAAAATCTAGAGCTGAAGAAATAAACTCCTAGAGAAAGCTGAAAGGGAGGAGG
CTGAAGTGGAGAAAAGTAGTGCTCATACCCAGGCCACCTGCTTTTGCAAGAAAAGTATGACAGTATGGTGCAAG
CCTTGAAGATGTTACTGCTCAATTTGAAGCTATAAAGCGTTAAGCAAGGAGGAGGAACTTTAGAAAAACAGCTGGAA
TCATCATACAGGAAAAAGCGCCAGGCTGGGAAAAATGCAGAGGATGTTGAGCATCAGATTTTGGCAACTGAGAGCT
CAATCAAGAAATATGTAAGGATGCTTCTAGATCTGCAGACCAAGTCAGCACTAAAGGAAACAGAAATTAAGAAATCAC
AGTTTCTTTTCTTCAAAAAATACTGATTGTCAGAACCACTCAAGCAACAGGAGGAGAACTTTAGAAAAACAGCTGGAA
GATGAAGAGGAAGAAAAGCTGAAAAAGAAAAATACAACAGCAGAAATTAAGTGAAGAAATTAACAAGTGGCGTCTCTCT
ATGAAGAACTATATAATAAAACAAACCTTTTCAGCTACAACCTAGATGCTTTTGAAGTAGAAAAACAGGCAATTGTTGAA
TGAAACATGGTGCAGCTCAGGAACAGCTAAATAAAATAAGAGATTATGCTAAATATTGGGTATCAGAAATTTGAAA
CAAAAAATCAAGCATGTTGTGAAGTTGAAAGATGAAATAGCCAACTCAATCGGAAGTATCAAACTCCGCTGTCAGC
TTGCTAAAAAAAACAAAGTGAGACAAACCTTCAAGAGGAATTGAATAAGTTCTAGGTATCAAACTTTGATCCTTC
AAAGGCTTTTCATCATGAAAGTAAAGAAAAATTTGCCCTGAAGACCCATTAAAGAGGCAATACAACTGTTACCGA
GCTCCTATGGAGTGTCAAGAAATCATGGAAGTAAACATCTGAGAAACCTGTTGAA
```

The NOV6c nucleic acid was identified on chromosome 5q33.2 and has 1944 of 1956 bases (99%) identical to a *Homo sapiens* intracellular hyaluronic acid binding protein (IHABP) mRNA (gb:GENBANK-ID:AF032862|acc:AF032862.1) (E = 0.0)

A disclosed NOV6c polypeptide (SEQ ID NO:18) encoded by SEQ ID NO:17 is 709 amino acid residues and is presented using the one-letter code in Table 6F. Signal P, Psort and/or Hydropathy results predict that NOV6c does not contain a signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500.

Table 6F. Encoded NOV6c protein sequence (SEQ ID NO:18).

MSFPPKAPLKRFPDPSGAPSPGAYDVKTLEVLKGPVSFQKSQRFKQKQESKQNLNVDKDTTLPASARKVKSSSEKIRVLLQERGA
QDRRIQDLETELEKMEARLNALREKTSLSANNATLEKQLIELTRTNELLKSKFSENDNQKNLRILSLMLKLRNKRETKMRGMM
AKQEGMEMKLVQTSRLEESQGKIAQLEBGLVSIIEKEKIDEKSETEKLLLEYIEEISCASDQVEKYKLDIAQLEENLKEKNDEILS
LKQSLSEENIVILSKQVEDLVNVCQQLLEKEKEDHVNRRNHNENLNAEMQNLKQKFIIEQQEREKLQKQELQIDSLQOEKELSSS
LHQKLCSPQEEVMVEKNLFEEELKQTLDELDKLQQKEEQAEERLVKQLEEEAKSRAEELKLLLEKLGKEAELEKSSAAHTQATLL
LQEKYDSMVQSLEDVTAQFEGYKALTASEIEDLKLENSSLQEKAAGKNAEDVQHQILATESNQEYVRMLLDLQTKSALKETE
IKEITVSFLQKITDLQNLKQOEEDFRKQLEDEEGRKAEKENTAEELTEEINKWRLLYEELYNKTKPFQLQDAFEVEKQALLNE
HGAAQEQNLKIRDSYAKLLGHQNLKQKI KHVVKLKDENSQKSEVSKLRQCLAKKKQSETKLQEBLNKVLGIKHFDPKAFHHES
KENFALKTPLKEGNTNCRAPMECQESWK

The NOV6c amino acid sequence has 706 of 724 amino acid residues (97%) identical to, and 706 of 724 amino acid residues (97%) similar to, a *Homo sapiens* 724 amino acid residue hyaluronan mediated motility receptor (intracellular hyaluronic acid binding protein) (receptor for hyaluronan-mediated motility) protein (ptnr:SWISSPROT-ACC:O75330) (E = 0.0).

NOV6c is expressed in at least the following tissues: Heart, Artery, Coronary Artery, Stomach, Liver, Appendix, Colon, Bone Marrow, Lymph node, Tonsils, Brain, Cervix, Mammary gland/Breast, Ovary, Placenta, Uterus, Prostate, Testis, Lung, Bronchus, Kidney Cortex, Retina and Epidermis. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

Possible small nucleotide polymorphisms (SNPs) found for NOV6a and NOV6c are listed in Tables 6G and 6H, respectively.

Table 6G: SNPs

Variant	Nucleotide Position	Base Change	Amino Acid Position	Base Change
13375250	23	C > A	Silent	N/A
13375251	51	G > T	5	Lys > Asn
13375231	52	G > A	6	Ala > Thr
13375252	89	C > T	18	Ala > Val
13375230	125	A > G	30	Glu > Gly
13375229	174	A > G	Silent	N/A
13375228	238	A > G	68	Lys > Glu
13375253	254	A > G	73	Glu > Gly
13375254	265	C > T	77	Arg > Cys
13375223	298	A > T	88	Arg > Trp
13374821	308	A > G	91	Gln > Arg
13375222	329	A > G	98	Glu > Gly
13375221	361	A > G	109	Arg > Gly
13375255	375	T > C	Silent	N/A
13375256	423	C > T	Silent	N/A
13375257	424	A > G	130	Arg > Gly
13375220	434	A > G	133	Glu > Gly

13375219	493	T > C	Silent	N/A
13375258	511	A > G	159	Arg > Gly
13375259	555	A > G	Silent	N/A
13375218	1090	A > G	352	Met > Val
13374820	1774	C > A	580	Leu > Ile
13375216	1909	C > T	625	His > Tyr

Table 6H: SNPs			
Consensus Position	Depth	Base Change	PAF
18	22	A > T	0.227
428	37	T > G	0.378
741	27	T > C	0.481
834	29	T > C	0.483
1750	37	C > T	0.351
1909	40	T > C	0.425
1952	41	G > A	0.073
2196	33	T > G	0.424

NOV6a - NOV6c are very closely homologous as is shown in the amino acid alignment in Table 6I.

Table 6I Amino Acid Alignment of NOV6a - NOV6c

	10	20	30	40	50	60	70
NOV6a	MSFPKAPLKR	FNDPSGCAPSPGAYDVKT	LEVLKGPVVSFOK	SQRFKQKQESKQNL	NVDKDTTLPASARKVK		
NOV6b	MSFPKAPLKR	FNDPSGCAPSPGAYDVKT	LEVLKGPVVSFOK	SQRFKQKQESKQNL	NVDKDTTLPASARKVK		
NOV6c	MSFPKAPLKR	FNDPSGCAPSPGAYDVKT	LEVLKGPVVSFOK	SQRFKQKQESKQNL	NVDKDTTLPASARKVK		
	80	90	100	110	120	130	140
NOV6a	SSSEKIRVLLQ	ERGAQDRRIQDLETELEKMEAR	LNAAALREKTSLSANNATLEKQ	LIELTRTNELLKSKFS			
NOV6b	SSSEKIRVLLQ	ERGAQDRRIQDLETELEKMEAR	LNAAALREKTSLSANNATLEKQ	LIELTRTNELLKSKFS			
NOV6c	SSSEKIRVLLQ	ERGAQDRRIQDLETELEKMEAR	LNAAALREKTSLSANNATLEKQ	LIELTRTNELLKSKFS			
	150	160	170	180	190	200	210
NOV6a	ENGNGOKNLR	ILSLELMKLRN	KRETKMRGMMAKQEGMEMKLQVTORS	LEESQGKIAOLEGKLV	SIEKEKID		
NOV6b	ENGNGOKNLR	ILSLELMKLRN	KRETKMRGMMAKQEGMEMKLQVTORS	LEESQGKIAOLEGKLV	SIEKEKID		
NOV6c	ENGNGOKNLR	ILSLELMKLRN	KRETKMRGMMAKQEGMEMKLQVTORS	LEESQGKIAOLEGKLV	SIEKEKID		
	220	230	240	250	260	270	280
NOV6a	EKSETEKLL	EYIEEISCASDQVEKYKLDIAQLEENL	KEKNDEILSLKQSL	EDNIVILSKQVEDLN	VKCOL		
NOV6b	EKSETEKLL	EYIEEISCASDQVEKYKLDIAQLEENL	KEKNDEILSLKQSL	EDNIVILSKQVEDLN	VKCOL		
NOV6c	EKSETEKLL	EYIEEISCASDQVEKYKLDIAQLEENL	KEKNDEILSLKQSL	EDNIVILSKQVEDLN	VKCOL		
	290	300	310	320	330	340	350
NOV6a	LETEKEDHV	NRRNREHNENLNAEMONL	EQKFILEQREHEKLQOKELQ	IDSLLQOEKELSSSLH	QKLC	SFQ	
NOV6b	LETEKEDHV	NRRNREHNENLNAEMONL	EQKFILEQREHEKLQOKELQ	IDSLLQOEKELSSSLH	QKLC	SFQ	
NOV6c	LETEKEDHV	NRRNREHNENLNAEMONL	EQKFILEQREHEKLQOKELQ	IDSLLQOEKELSSSLH	QKLC	SFQ	
	360	370	380	390	400	410	420
NOV6a	EMVKEKNL	FEEELKQTLDEL	DKLQKEEQAE	RLVKQLEEEAKSRAEEL	KLLEEK	KLKGKAELEKSSAAHT	
NOV6b	EMVKEKNL	FEEELKQTLDEL	DKLQKEEQAE	RLVKQLEEEAKSRAEEL	KLLEEK	KLKGKAELEKSSAAHT	
NOV6c	EMVKEKNL	FEEELKQTLDEL	DKLQKEEQAE	RLVKQLEEEAKSRAEEL	KLLEEK	KLKGKAELEKSSAAHT	
	430	440	450	460	470	480	490

- [illegible]

500 510 520 530 540 550 560

NOV6a NAEDVOHOILATESSNOBYVRMLLDLOTSAKETEIKEITVSFLOKITDLCNOLKQEEEDFRKOLEDEE

```

gi|7108351|NAEDVQHQILATESSNQEYVRMLLDLQTKSALKETEIKEITVSFLQKITDLOLNOLKQOEEDFRKQLEDEE
gi|1657698|NAEDVQHQILATESSNQEYVRMLLDLQTKSALKETEIKEITVSFLQKITDLOLNOLKQOEEDFRKQLEDEE
gi|2135413|NAEDVQHQILATESSNQEYVRMLLDLQTKSALKETEIKEITVSFLQKITDLOLNOLKQOEEDFRKQLEDEE
gi|7108349|NAEDVQHQILATESSNQEYVRMLLDLQTKSALKETEIKEITVSFLQKITDLOLNOLKQOEEDFRKQLEDEE
gi|4580681|RVEDVQOQILATESSNQEYAKVQDLQNSSTLKEAEIKEITSSYLEKITDLOLNOLKQOEEDFRKQLEDEE

      570      580      590      600      610      620      630
NOV6a      GRKAEKENTTAELTEEINKWRLLYEELYNKTKPPQLQDAFEVEKQALLNEHGAAQEQLNKIRDSYAKLL
gi|7108351|GRKAEKENTTAELTEEINKWRLLYEELYNKTKPPQLQDAFEVEKQALLNEHGAAQEQLNKIRDSYAKLL
gi|1657698|GRKAEKENTTAELTEEINKWRLLYEELYNKTKPPQLQDAFEVEKQALLNEHGAAQEQLNKIRDSYAKLL
gi|2135413|GRKAEKENTTAELTEEINKWRLLYEELYNKTKPPQLQDAFEVEKQALLNEHGAAQEQLNKIRDSYAKLL
gi|7108349|GRKAEKENTTAELTEEINKWRLLYEELYNKTKPPQLQDAFEVEKQALLNEHGAAQEQLNKIRDSYAKLL
gi|4580681|AKMTEKETAVTELTMEINKWRLLYEELDKTKPPQQLQDAFEKQALLNEHGATQEQLSKIRDSYAKLL

      640      650      660      670      680      690      700
NOV6a      GHQNLKQKIKHVVKLKDENSQLKSEVSKLRCQLAKKKQSETKLOEELNKVLGIKHFDPSKAFHHESKENF
gi|7108351|GHQNLKQKIKHVVKLKDENSQLKSEVSKLRCQLAKKKQSETKLOEELNKVLGIKHFDPSKAFHHESKENF
gi|1657698|GHQNLKQKIKHVVKLKDENSQLKSEVSKLRCQLAKKKQSETKLOEELNKVLGIKHFDPSKAFHHESKENF
gi|2135413|GHQNLKQKIKHVVKLKDENSQLKSEVSKLRCQLAKKKQSETKLOEELNKVLGIKHFDPSKAFHHESKENF
gi|7108349|GHQNLKQKIKHVVKLKDENSQLKSEVSKLRCQLAKKKQSETKLOEELNKVLGIKHFDPSKAFHHESKENF
gi|4580681|GHQNLKQKIKHVVKLKDENSQLKSEVSKLRCQLAKKKQSETKLOEELNKVLGIKHFDPSKAFHHESKENF

      710      720
NOV6a      ALKTPPLKEGNTNCRYAPMECCQESWK
gi|7108351|ALKTPPLKEGNTNCRYAPMECCQESWK
gi|1657698|ALKTPPLKEGNTNCRYAPMECCQESWK
gi|2135413|ALKTPPLKEGNTNCRYAPMECCQESWK
gi|7108349|ALKTPPLKEGNTNCRYAPMECCQESWK
gi|4580681|TLKTPPLKEGNTNCRYAPMECCQESWK

```

Hyaluronan is a large glycosaminoglycan that is ubiquitous in the extracellular matrix and whose synthesis has been linked to cell migration, growth and transformation. It interacts with cell surfaces via specific protein receptors, receptor for hyaluronic acid mediated

5 motility, that mediate many biological effects. Hardwick et al. (1992; *J Cell Biol* 117:1343-50) cloned a hyaluronan receptor cDNA from mouse 3T3 cells. The 2.9-kb cDNA codes for a predicted 477-amino acid protein, which they designated RHAMM. Antibodies directed against the protein blocked locomotion of cells induced by expression of a mutant H-ras. Savani et al. (1995; *J Clin Invest* 95:1158-68) showed that RHAMM is upregulated in

10 response to wound healing. When hyaluronan binds to RHAMM the phosphorylation of a number of proteins, including the focal adhesion kinase pp125-FAK, occurs (Hall et al., 1994; *J Cell Biol* 126:575-88). The latter is a necessary step for disassembly of focal contacts and subsequent motility. Entwistle et al. (1995; *Gene* 163: 233-8) showed that the mouse gene contains at least 14 exons spanning greater than 15 kb and can produce alternatively spliced

15 mRNAs, one of which is transforming (Hall et al., 1995; *Cell* 82:19-28), similar to the hyaluronan receptor CD44 (107269). Spicer et al. (1995; *Genomics* 30:115-7) used interspecific backcross analysis to map the mouse gene to chromosome 11 within a region of synteny to human chromosome 5q23-q35. They used somatic cell hybrid DNAs and a radiation hybrid panel to confirm the distal 5q map location (5q33.2-qter) of the HMMR gene

20 in human. The map position of the human RHAMM gene places it in a region comparatively

TCAGCAAGGACGCTATTAATGCTGAGACTGTGCTGGTACTGGTGAATGCTGTTACTTCAAGGCCAAATGGGAAACATAC
TTTGACCATGAAAACACGGTGGATGCACCTTTCTGTCTAAATCAGAATGAAAACAAGAGTGTGAAGATGATGACGCAAAA
AGGCCCTCTACAGAATTGGCTTCATAGAGGAGGTGAAGGCACAGATCCTGGAAATGAGGTACACCAAGGGGAAGCTCAGCA
TGTTTCGTGCTGCTGCCATCTCACTCTAAAGATAACCTGAAGGGTATCACCTATGAAAAAATGGTGGCCTGGAGCAGCTCA
GAAAACATGTCAGAAGAATCGGTGGTCTGTCTTCCCCGGTTCACCTGGAAGACAGCTATGATCTCAATTCATTTT
ACAAGACATGGGCATTACGGATATCTTTGATGAAACGAGGGTGTCTTACTGGAATCTCTCCAAGTCCCAATTTGTACT
TGTCAAAAATTATCCACAAAACCTTTGTGGAGGTGGATGAAACCGTACCCAGGCAGCTGCAGCCACTGGGGCTGTGTGTC
TCGGAAAGGTCACTACGATCTTGGGTGGAGTTTAATGCCAACCCCTTTCTCTTTTCATTAGACACAACAAACCCCA
AACCATTCTCTTTATGGCAGGGTCTGCTCTCTCTTAAAGGGGAGCAGTGTCTAGTACTTTGGAGCTGGAGGAAAA

The disclosed NOV7 nucleic acid sequence, localized to chromosome 18, has 258 of 408 bases (63%) identical to a *Homo sapiens* cytoplasmic antiproteinase 3 (CAP3) mRNA (gb:GENBANK-ID:HUMCAP3A|acc:L40378.1) ($E = 3.6e^{-41}$).

- 5 A disclosed NOV7a polypeptide (SEQ ID NO:20) encoded by SEQ ID NO:19 is 365 amino acid residues and is presented using the one-letter amino acid code in Table 7B. Signal P, Psort and/or Hydropathy results predict that NOV7 does not contain a signal peptide and is likely to be localized to the nucleus with a certainty of 0.6000 and to the microbody (peroxisome) with a certainty of 0.5439.

Table 7B. Encoded NOV7 protein sequence (SEQ ID NO:20).

MDSLVTANTKFCFDLQEIIGKDDRHKNIFFSPLSLAALGMVRLGARSDSAHQIDEVRSLLNNEGLVSCYFGQLLSKLDL
IKTDYTLSIANRLYGESSLGDKSETLSQKKKKKIIYTNAPDTIHTQDILWDLFLGKIKELFSKDAINAETVLVLNAVYF
KAKWETYFDHENTVDAPFCLNQENKSVKMMTQKGLYRIGFIEVKAQILEMRYTKGKLSMFVLLPSHSDNLKGIYK
MVAWSSSENMSSESVLSFPRFTLEDSYDLNSILQDMGITDIFDETADLTGISPSPNLYLSKIIHKTFVEVDENGTQAA
AATGAVVSESLRSWVEFNANHPFLFFIRHNKTQTILFYGRVCSF

10

The NOV7 amino acid sequence has 174 of 378 amino acid residues (46%) identical to, and 242 of 378 amino acid residues (64%) similar to the *Bos taurus* 378 amino acid residue serine proteinase inhibitor B-43 protein (ptnr:SWISSPROT-ACC:O02739) ($E = 8.5e^{-79}$).

Possible small nucleotide polymorphisms (SNPs) found for NOV7 are listed in

- 15 Table 7C.

Table 7C: SNPs				
Variant	Nucleotide Position	Base Change	Amino Acid Position	Base Change
13376217	236	C > T	Silent	N/A
13376218	240	A > G	61	Asn > Asp
13376220	620	T > C	Silent	N/A
13376221	634	C > T	192	Thr > Met

NOV7 has homology to the amino acid sequence shown in the BLASTP data listed in Table 7D.

Table 7D. BLAST results for NOV7

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
---------------------------	-------------------	----------------	-----------------	------------------	--------


```

gi|12843557|EDNVNSLQELKKINHEKLLAWSSSENLSKPPVAISFPQPNLEDSYDLKSILODMGILKQVDETRADLIG
gi|6685780|-SESTDLNTVEKALTYEKFLAWTKPDVMDDEEVEVFLPRFTLEBSYDMEEFLQELGMDAFETRADFSG
gi|12834891|EDSTGLKKEKQITLEKLLBWKRENLEFIDVHVKLPRFKIEBSYTNLSNLGRLEGVODLFSSSKADLSG
gi|12843390|EDSTGLKKEKQITLEKLLBWKRENLEFIDVHVKLPRFKIEBSYTNLSNLGRLEGVODLFSSSKADLSG

          360      370      380      390      400      410      420
NOV7      . . . . .
gi|16163747|ISPSPNLYLSKIIHKTFVEVDENGTA AAAATGAVVSERSLRSWVEFNANHPFLFFIRHNKTOTILFYGRV
gi|12843557|ISPSPNLYLSKIIHKTFVEVDENGTA AAAATGAVVSERSLRSWVEFNANHPFLFFIRHNKTOTILFYGRV
gi|6685780|ISPSPNLYLSKIIHKTFVEVDENGTA AAAATGAVVSERSLRSWVEFNANHPFLFFIRHNKTOTILFYGRV
gi|12834891|ISPSPNLYLSKIIHKTFVEVDENGTA AAAATGAVVSERSLRSWVEFNANHPFLFFIRHNKTOTILFYGRV
gi|12843390|ISPSPNLYLSKIIHKTFVEVDENGTA AAAATGAVVSERSLRSWVEFNANHPFLFFIRHNKTOTILFYGRV

NOV7      . . . . .
gi|16163747|CSF
gi|12843557|CSF
gi|6685780|YCF
gi|12834891|CSF
gi|12843390|CSF

```

Tables 7F and 7G list the domain description from DOMAIN analysis results against NOV7. This indicates that the NOV7 sequence has properties similar to those of other proteins known to contain these domains.

Table 7F Domain Analysis of NOV7

gnl|Pfam|pfam00079, serpin, Serpin (S protease inhibitor). Structure is a multi-domain fold containing a bundle of helices and a beta sandwich. (SEQ ID NO:87)
Length = 377 residues, 98.4% aligned
Score = 280 bits (716), Expect = 1e-76

5

NOV7:	3	SLVTANTKFCFDLQFEIGKDDRHNKIFFSPLSLSAALGMVRLGARSDSAHQIDEVRSLLN	62
		+ ++ + + + + + + + +	
00079:	7	KLASANADFAFSLYKELVEQNPDKNIFFSPVSISSALAMLSLGAKGNTATQILEVLGFNL	66
NOV7:	63	ESGLVSCY---FGQLLSKLDRIKTDYTLSIANRLYGESSLG-----DKSETLSQKKKKK	113
		+ ++ + + + + + +	
00079:	67	TETSEAEIHQGFQHLQLQELNRPDTGLQLTTGNALFVDKSLKLLDEFLEDSKRLYQSEVFS	126
NOV7:	114	IIYTNAFDTIH-TQDILWDLFLGKIKELFSKDAINAETVLVLVNAVYFKAKWETYFDHEN	172
		+ + + + + + + + + + +	
00079:	127	VDFSDEEAKKQINDWVEKKTQGKIKDLLK--DLSDTDVLVLVNYIYFKGKWKPPDPDEL	184
NOV7:	173	TVDAPFCLNQENKSVKMMTQKGLYRIGFIEEVKAQILEMRYTKGKLSMFVLLPSHSDN	232
		+ + + + + + + +	
00079:	185	TEEDFHVDKKTTVKVPMMNQLGTFFYYRDEELNCKVLELPYKGNATSMFLFDPDEVGKL	244
NOV7:	233	--LKGITYEKMVAWSSSENEMSESVVLSFPFRFTLEDSYDLNSILODMGITDIFDETRAD	289
		++ +	
00079:	245	EQVEAALSPETLRKW--LENMEPREVELYLPKFSIEGTYDLKDVLA KLGITDLFSN-QAD	301
NOV7:	290	LTGISPSPNLYLSKIIHKTFVEVDENGTA AAAATGAVVSERSLRSWVEFNANHPFLFFIR	349
		+ + + +	
00079:	302	LSGISEDEDLKVS KAVHKAVLEVDEEGTEAAAATGAIIVPRSLPPELEFTADRPFLLIY	361
NOV7:	350	HNKTQTILFYGRVCSP	365
		+ + + +	
00079:	362	DDPTGSILFMGKV VNP	377

Table 7G Domain Analysis of NOV7

gnl|Smart|smart00093, SERPIN, Serine Proteinase Inhibitors (SEQ ID NO:88)
 Length = 360 residues, 100.0% aligned
 Score = 275 bits (702), Expect = 4e-75

```

NOV7: 13  FDLFQEIGKDDRHKNIFFSPLSLSAALGMVRLGARSDSAHQIDEVRSLNNESEGLVSC--- 69
      |||++|+ |+ |||||+|+|+| |+ |||+ +| || || | +
00093: 1  FDLYKELAKESPDKNIFFSPVSISSALAMLSLGAKGSTATQILEVLGFNLTETSEADIHQ 60

NOV7: 70  YFQQLLSKLDRIKTDYTLSIANRLYGESSLG-----DKSETLSQKKKKKIIYTNAFDTI 123
      | || | |+| | || |+ + || + + | + + + + + +
00093: 61  GFQHLHLHLNRPDNKLQLKTANALFVDKSLKLLDSFLEDVKKLYGAEVQSVDFSDPAEEA 120

NOV7: 124 HTQDILW--DLFLGKIKELFSKDAINAETVLVLVNAVYFKAKWETYFDHENTVDAPFCLN 181
      | | | |||+| | ++ +| |||||+||| ||+| || ||| + | ++
00093: 121 KKQINDWVKKKTQGKIKDLLS--DLDPDTRLVLVNAIYFKGKWKTTPDPENTREEDFYVD 178

NOV7: 182 QNENKSVKMMTQKG-LYRIGFIEEVKAQILEMRYTKGKLSMFVLLPSHSK--DNLKGITY 238
      + | | | |+| +| | ||| +|+|+| | | | | ++| | +|
00093: 179 ETTTVKVPMMSTQGRTRYGRDEELNCQVLELPY-KGNASMLIILPDEGGLETVEKALTP 237

NOV7: 239 EKMVAWSSSENSESVLSFPRFTLEDSDYLNLSILQDMGITDIFDETRADLTGISPSPN 298
      | + | ++++++ || | |+| || ||||| +|+ +|+|||+| |||+||| +
00093: 238 ETLKKW--TKSLTKRSVELYLPKFKLEISYDLKDVLEKLGITDLFSNK-ADLSGISEDKD 294

NOV7: 299 LYLSKIIHKTFVEVDENGTAQAAAATGAVVSERSLRSWVEFNANHPFLFFIRHNKTQTILF 358
      | +|+|+|| |+|+| ||+||| || ++ || | | ||| || | | +|||
00093: 295 LKVSQVVKHKAFLVNEEGTEAAAATGVIIIVPRSLPP-PEFKANRPFLIRDNPTGSILF 353

NOV7: 359 YGRVCSP 365
      |+| +|
00093: 354 MGKVVP 360
  
```

Serpins are protease inhibitors that have applications to tissue regeneration and the treatment of tumors.

- 5 Schneider et al. (Proc Natl Acad Sci U S A 92(8):3147-51,1995) demonstrated that 18q21.3 contains a cluster of serine proteinase inhibitors, or 'serpins,' including a tandem duplication of the squamous cell carcinoma antigen (SCCA) gene and 2 other members of the ovalbumin family of serine proteinase inhibitors, plasminogen activator inhibitor type 2 (173390) and maspin (protease inhibitor-5; 154790). Schneider et al. (1995) presented
- 10 evidence that the neutral and acidic forms of SCCA are encoded by SCCA1 (600517) and SCCA2, respectively.

- 15 Barnes and Worrall (FEBS Lett. 373(1):61-5, 1995) described the cloning of a member of the serpin family of serine protease inhibitors by degenerate PCR and screening of a HeLa cell cDNA library. The isolated cDNA encodes a 390-amino acid protein, designated leupin by them, that is 91.8% identical to SCCA1. The authors stated that the reactive site of leupin differs from SCCA1 by the presence of a leucine residue rather than a serine at the P(1) position within the loop region that acts as a pseudosubstrate for the target protease. Barnes

and Worrall (1995) speculated that leupin may be a cysteine protease inhibitor. Schick et al. (J Biol Chem. 272(3):1849-55, 1997) demonstrated that SCCA2 inhibits the chymotrypsin-like proteinases cathepsin G (116830) and mast cell chymase (118938) in vitro. SCCA2 was ineffective against papain-like cysteine proteinases, which have been shown to be inhibited by SCCA1.

The mammalian liver has an extraordinary capacity for regeneration. In the rat, the liver regenerates the most of its original mass within several days following hepatectomy; regeneration is virtually complete by 2 weeks after surgery. New et al. (Biochem Biophys Res Commun.223(2):404-12, 1996) isolated a gene encoding a plasma protein by constructing and screening a cDNA library with RNA isolated from liver at 48 hours after 70 to 90% hepatectomy. New et al. (1996) stated that the expression of acute phase inflammatory proteins should be substantially diminished, thereby reducing the 'background' and facilitating the identification of genes associated with regeneration. They identified several clones that were upregulated in the regenerating liver. They isolated 1 clone, termed 'regeneration-associated serpin-1' (RASP1), that was expressed in normal liver but was upregulated approximately 3- to 4-fold by 48 hours after hepatectomy. DNA sequence analysis showed that the RASP1 gene encodes a novel 436-amino acid secreted protein. Moderate homology was found with several members of the serpin family of serine-protease inhibitors. The 1.7-kb RASP1 mRNA was highly expressed in rat liver, but not in brain, heart, kidney, lung, testis, or spleen. It was found in normal and hepatectomy rat plasma

The above defined information for NOV7 suggests that this NOV7 protein may function as a member of a Serpin protein family. Therefore, the NOV7 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the NOV7 compositions of the present invention will have efficacy for treatment of patients suffering from liver toxicity, cancer, metabolic diseases, inflammation, CNS disorders and other diseases, disorders and conditions of the like. The NOV7 nucleic acid encoding Serpin-like protein, and the Serpin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV8

NOV8 includes seven novel B7 Family-like proteins disclosed below. The disclosed proteins have been named NOV8a – NOV8g.

NOV8a

A disclosed NOV8a nucleic acid of 1590 nucleotides (also referred to CG50309-01) encoding a novel B7 family-like protein is shown in Table 8A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 16-18 and ending with a TGA codon at nucleotides 1411-1413. Putitive untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 8A, and the start and stop codons are in bold letters.

Table 8A. NOV8a Nucleotide Sequence (SEQ ID NO:21)

CAGCTCCCGGGCACC**ATG**CGAACCGCCCGAGCCTCCGCGCTGCGTCTGCCTGTGCTCGCCCGCATCTGGACCTGGC
GCGCTACCTGACAGTCAACATTGAGCCTCTCCCCCTGTGGTGGCTGGAGACGCCGTGACTTTGAAGTGAACCTCAAGA
CAGATGGGGCGCATGCCGGAGATCGTGTGGTACCGGGTGACGGATGGTGGCACCATCAAGCAAAAGATCTTCACTCTCGAC
GCCATGTTCTCCACCACTACTCACACATGGAGAATACCGCAAGCGAGAGGACCTGGTGTACCACTGCTGAGGCT
GCCCCGAGGTCCGGATCTCAGACAATGGTCCCTATGAGTGCCATGTGGGCATCTACGACCGGCCACAGGGAGAAGGTGG
TCCTGGCATCAGGCAACATCTTCTCAACGTCATGGCTCCTCCACCTCCATTGAAGTGGTGGCTGTGACACACCAGCC
CCCTTCAGCCGCTACCAAGCCCAAGACTTCACGCTGGTCTGCATCGTGTCTGGAGGAAAACAGCACCCATGGTTTATTT
CAACGAGATGGGGAACCAATCGACGCACTGCCCTATCAGAGCCACAGCTGCGAGCTCCGGCCCCCTACAGGACAGCA
GGCCCTTCCGACGCTTCTGCTGGACCTGGATGACACCAAGATGCAGAAGTCACTGTCCCTCCTGGACGCCGAGAACCGG
GGTGGGCGACCTTACACGGAGCGCCCTCCCGTGGCTGACCCAGATCCCAACATCCTCCTCCAGCCAAACACAGAGAA
CATACCAGAGACGGTCGTGAGCCGTGAGTTTCCCGCTGGGTCCACAGCGCCGAGCCCACTACTTCTGCGCCACAGCC
GCACCCCGAGCAGTGACGGCACTGTGGAAGTACGTGCCCTGCTCACCTGGACCTCAACCCACAGATCGACAACGAGGCC
CTCTTCAGCTGCGAGGTCAAGCACCCAGCTCTGTGATGCCATGCGGGCAGAGGTACGCGCGTTGCCCCCAAAGGACC
CAAAATTGTGATGACGCCAGCAGAGCCCGGTAGGGGACACAGTGAAGATTCTGGTCCATGGGTTTCAGAACGAAGTCT
TCCCGGAGCCCATGTTACGTGGACCGGGTTGGAGCCGCTCCTGGACCGCAGCGCTGAGTTCGACGGGAAGGAGCTG
GTGCTGGAGCGGGTTCCCGCCGAGCTCAATGGTCCATGTATCGCTGCACCGCCAGAACCCACTGGGCTCCACCGACAC
GCACACCCGGCTCATCGTGTGTTGAAAACCCAAATATCCCAAGAGGAACGGAGGACTCTAATGGTTCATTGGCCCCACTG
GTGCCCGGCTCACCTTGGTGTCTGCCCTGACAGTGAATCTGGAGCTGACGTGAAGGCACCCGCCCGGCCATCCATCAG
GCACGTGACATCTCCAGACCGGTTTTCATTTCTTTCTAAACTATTTCCAGTCTTGTCTTAGTCTCTTTCCATCTGTGT
CTTGGCTTCTTCAGTCGGTTAATTAAACAAACAGAACAAATTTCCCAAAAAAAAAAAAAAAAAAAAAA

The disclosed NOV8a nucleic acid sequence, localized to chromosome 1, has 1535 of 1595 bases (96%) identical to a *Macaca fascicularis* brain cDNA, clone:QccE-13927 (gb:GENBANK-ID:AB046009|acc:AB046009.1) (E = 0.0).

A disclosed NOV8a polypeptide (SEQ ID NO:22) encoded by SEQ ID NO:21 is 465 amino acid residues and is presented using the one-letter amino acid code in Table 8B. Signal P, Psort and/or Hydropathy results predict that NOV8a contains a signal peptide and is likely to be localized extracellularly with a certainty of 0.6902. The most likely cleavage site for a NOV8a peptide is between amino acids 37 and 38, at: VVA-GD.

Table 8B. Encoded NOV8a protein sequence (SEQ ID NO:22).

MRTAPSLRRVCVLLLAAILDLARYLTVNIPLPPVVGDAVTLKCNFKTDGRMREIVWYRVTDGGTIKQKIFTFDAMFST
NYSHMENYRKREDLVYQSTVRLPEVRI SDNGPYECHVG IYDRATREKVVLASGNI FLNVMAPPTSIEVVAADTPAPFSRY
QAQNFLLVCI VSGGK PAMVYFKRDGEPIDAVPLSEPPAASSGPLQDSRPFRLSLDLDDTKMQKSLSLDAENRGRPY
TERPSRGLTPDPNILLQPTTENI PETVVSREFPRVHSAEPTYFLRHSRTPSSDGTVEVRALLTWTLPQIDNEALFSC
VKHPALSMPMRAEVTVPAPKGP KIVMTPSRARVGDVRLVHGFQNEVFPEPMFTWTRVGSRLLDGSAEFDGKELVLERV
PAELNGSMYRCTAQNLGSTDHTRLIVFENPNIPRGTEDSNGSIGPTGARLTLVLALTIVILELT

The NOV8a amino acid sequence has 396 of 404 amino acid residues (98%) identical to, and 397 of 404 amino acid residues (98%) similar to, a *Macaca fascicularis* 404 amino acid residue protein (ptnr:SPTREMBL-ACC:Q9N0C1) (E = 1.8e⁻²¹¹).

NOV8a is expressed in at least the following tissues: Brain, Heart, Thalamus, Lung, Pancreas, Prostate and Whole Organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, NOV8a is predicted to be expressed in brain tissues because of the expression pattern of a closely related *Macaca fascicularis* brain cDNA homolog, clone:QccE-13927 (gb:GENBANK-ID:AB046009|acc: AB046009.1).

NOV8b

A disclosed NOV8b nucleic acid of 1593 nucleotides (also referred to CG50309-02) encoding a novel B7 family-like protein is shown in Table 8C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 16-18 and ending with a TGA codon at nucleotides 1414-1416. Putitive untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 8C, and the start and stop codons are in bold letters.

Table 8C. NOV8b Nucleotide Sequence (SEQ ID NO:23)

CAGCTCCCGGGACCATGCGAACCGCCCGAGCCTCCGCGCTGCGTCTGCCTGCTGCTCGCCGCGATCCTGGACCTGGC
GCGCGGCTACCTGACAGTCAACATTGAGCCTCTCCCCCTGTGGTGGCTGGAGACGCCGTGACTTTGAAGTGTAACCTCA
AGACAGATGGGCGCATGCGGAGATCGTGTGTACCGGTGACGGATGGTGGCACCATCAAGCAAAAGATCTTCACCTTC
GACGCCATGTTCTCCACCACTACTCACACATGGAGAATACCGCAAGCGAGAGGACCTGGTGTACCACTGTGAG
GCTGCCCCGAGGTCCGGATCTCAGACAATGGTCCCTATGAGTGCCATGTGGGCATCTACGACCGGCCACAGGGAGAAGG
TGGTCTGGCATCAGGCAACATCTTCCTCAACGTCTGCTCCTCCACCTCCATTTGAAGTGGTGGCTGTGACACACCA
GCCCCCTTCAGCCGCTACCAAGCCCGAAGCTTACGCTGGTCTGCATCGTGTCTGGAGAAAACAGCACCCATGGTTTA
TTTCAAACGAGATGGGGAACCAATCGACGAGTGCCCTATCAGAGCCACAGCTGCGAGCTCCGGCCCCCTACAGGACA
GCAGGCCCTTCGCGAGCCTTCTGCTGGACCTGGATGACACCAAGATGCAGAACTCACTGTCCCTCCTGGACGCCGAGAAC
CGGGGTGGGCGACCCCTACAGGAGCGCCCTCCCGTGGCCTGACCCAGATCCCAACATCCTCTCCAGCCAACACAGA
GAACATACAGAGACGGTCTGTAGCCGTGAGTTTCCCCGCTGGGTCCACAGCGCCGAGCCACCTACTTCTGCGCCACA
GCCGCACCCCGAGCAGTGACGGCACTGTGGAAGTACGTGCCCTGCTCACCTGGACCCCTCAACCCACAGATCGACAACGAG
GCCCTCTTCAGCTGCGAGGTCAAGCACCCAGCTCTGTGATGCCATGCCATGCCGCGAGAGGTACGCGCGTTGCCCCAAAGG
ACCCAAATTTGTGATGACGCCAGCAGAGCCGGGTAGGGGACACAGTGAGGATTTGGTCCATGGGTTTCAGAACGAAG
TCTTCCCGAGGCCATGTTTACGTGGACGCGGGTTGGGAGCCGCTCTGGACGGCAGCGCTGAGTTGACGGGAAGGAG
CTGGTGTGGAGCGGGTTCCGCGGAGCTCAATGGCTCCATGTATCGCTGCACCGCCAGAACCCACTGGGCTCCACCGA
CACGCACACCCGGCTCATCGTGTTTGAAACCCAAATATCCCAAGAGGAACGGAGACTCTAATGGTTCCATTGGCCCCA
CTGTGCCCCGGCTCACCTTGGTGTGCGCCTGACAGTGATTCTGGAGCTGACGTGAAGGCACCCCGCCCGGCCACTCCAT
CAGGCACTGACATCTCCACGACCGGTTTTCATTTCTTTCTAAACTATTTCCAGTCTTGTCTTAGTCTCTTTCCATCTG
TGTCTTGGCTTCTTCAGTCGGTTTAATTAACAAACAGAACAAATTTCCCCAAAAAAAAAAAAAAAAAAAA

The disclosed NOV8b nucleic acid sequence, localized to chromosome 1, has 1536 of 1595 bases (96%) identical to a *Macaca fascicularis* brain cDNA, clone:QccE-13927 (gb:GENBANK-ID:AB046009|acc:AB046009.1) (E = 0.0).

A disclosed NOV8b polypeptide (SEQ ID NO:24) encoded by SEQ ID NO:23 is 466 amino acid residues and is presented using the one-letter amino acid code in Table 8D. Signal P, Psort and/or Hydropathy results predict that NOV8b contains a signal peptide and is likely to be localized extracellularly with a certainty of 0.6233. The most likely cleavage site for a NOV8b peptide is between amino acids 24 and 25, at: ARG-YL.

Table 8D. Encoded NOV8b protein sequence (SEQ ID NO:24).

MRTAPSLRRCVCLLLAAILDLARGYLTVNIEPLPPVAGDAVTLKCNFKTDGRMREIVWYRVTDGGTIKQKIFTDFDAMPS
 TNYSHMENYRKREDLVYQSTVRLPEVRI SDNGPYECHVGIYDRATREKVVLASGNI FLNVMA PPTSIEVVAADTPAPFSR
 YQAQNTFLVCIVSGGKPAPMVYFKRDGEPI DAVPLSEPPAASSGFLQDSRPFRLDLDLDDTKMQKSLSLDAENRGGRP
 YTERPSRGLTPDPNILLQPTTENIPETVVSREFFRWVHSAEPTYFLRHSRTPSSDGTVEVRALLTWLNPQIDNEALFSC
 EVKHPALSMPMRAEVPVAPKGPKI VMTPSRARVGDVRLLVHGFQNEVFPPEPMFTWTRVGSRLLDGSAEFDGKELVLER
 VPAELNGSMYRCTAQNPLGSTDTHTRLIVFENPNI PRGTEDSNGSIGPTGARLTLVLALTIVILELT

The NOV8b amino acid sequence has 397 of 404 amino acid residues (98%) identical to, and 398 of 404 amino acid residues (98%) similar to, a *Macaca fascicularis* 404 amino acid residue protein (ptnr:SPTREMBL-ACC:Q9N0C1) ($E = 5.9e^{-213}$).

NOV8b is expressed in at least the following tissues: Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, NOV8b is predicted to be expressed in brain tissues because of the expression pattern of a closely related *Macaca fascicularis* brain cDNA homolog, clone:QccE-13927 (gb:GENBANK-ID:AB046009|acc: AB046009.1).

NOV8c

A disclosed NOV8c nucleic acid of 1407 nucleotides (also referred to CG50309-03) encoding a novel B7 family-like protein is shown in Table 8E. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 1402-1404. A putitive untranslated region downstream from the termination codon is underlined in Table 8E, and the start and stop codons are in bold letters.

Table 8E. NOV8c Nucleotide Sequence (SEQ ID NO:25)

ATGCGAACCGCCCGAGCCTCCGCCGCTGCGTCTGCCTGCTGCTCGCCGCGATCCTGGACCTGGCGCGGGCTACCTGAC
 AGTCAACATTGAGCCTCTCCCCCTGTGGTGGCTGGAGACGCCGTGACTTTGAAGTGTAACCTCAAGACAGATGGGCGCA
 TGCGGGAGATCGTGTGGTACCGGGTGACGGATGGTGGCACCATCAAGCAAAAGATCTTCACTTCGACGCCATGTTCTCC
 ACCAACTACTCACACATGGAGAACTACCGCAAGCGAGAGGACCTGGTGTACCACTCCACTGTGAGGCTGCCCGAGGTCGG
 GATCTCAGACAATGGTCCCTATGAGTGCCATGTGGGCATCTACGACCGCGCCACCAGGGAGAAGGTGGTCCCTGGCATCAG
 GCAACATCTTCTCAACGTCATGGCTCCTCCCACTCCATTGAAGTGGTGGCTGCTGACACACCAGCCCTTTCAGCCGC
 TACCAAGCCCGAGAACTTCAAGCTGGTCTGCATCGTGTCTGGAGGAAAACAGCACCCATGGTTTATTTCAAACGAGATGG
 GGAACCAATCGACGAGTGGCCCTATCAGAGCCACCAGCTGCGAGCTCCGGCCCCCTACAGGACAGCAGGCCCTTCCGCA
 GCCTTCTGCACCGTGACCTGGATGACACCAAGATGCAGAAGTCACTGTCCCTCCTGGACGCCGAGAACCGGGTGGCGCA
 CCCTACACGGAGCGCCCTCCCGTGGCCTGACCCAGATCCCAACATCCTCCTCCAGCCAACCAAGAGAGAACATACCAGA
 GACGGTCGTGAGCCGTGAGTTTCCCGCTGGGTCCACAGCGCCGAGCCCACTACTCCTGCGCCACAGCCCGACCCCGA
 GCAGTGACGGCACTGTGGAAGTACGTGCCCTGCTCACTGGACCTCAACCCACAGATCGACAACGAGGCCCTTTCAGC
 TGCAGAGTCAAGCACCCAGCTCTGTGATGCCCATGCAGGCAGAGGTACGCTGGTTGCCCCCAAGGACCCAAAATTGT
 GATGACGCCCGAGCAGCCCGGTAGGGGACACAGTGAAGATTCTGGTCCATGGGTTTCAGAACGAAGTCTTCCCGGAGC
 CCATGTTACAGTGGACGCGGGTGGGAGCCGCCTCCTGGACGGCAGCGCTGAGTTCGACGGGAAGGAGCTGGTCTGGAG

CGGGTTCGCGCCGAGCTCAATGGCTCCATGTATCGCTGCACCGCCAGAACCCACTGGGCTCCACCGACACGCACACCCG
GCTCATCGTGTGTTGAAACCCAAATATCCCAAGAGGAACGGAGGACTTAATGGTTCATTGGCCCCACTGGTGCCCGGC
TCACCTTGGTGCTCGCCCTGACAGTGATTCTGGAGCTGACGTGAAGG

The disclosed NOV8c nucleic acid sequence, localized to chromosome 1, has 1363 of 1407 bases (96%) identical to a *Macaca fascicularis* brain cDNA, clone:QccE-13927 (gb:GENBANK-ID:AB046009|acc:AB046009.1) ($E = 3.5e^{-294}$).

A disclosed NOV8c polypeptide (SEQ ID NO:26) encoded by SEQ ID NO:25 is 467 amino acid residues and is presented using the one-letter amino acid code in Table 8F. Signal P, Psort and/or Hydropathy results predict that NOV8c contains a signal peptide and is likely to be localized extracellularly with a certainty of 0.6233. The most likely cleavage site for a NOV8c peptide is between amino acids 24 and 25, at: ARG-YL.

Table 8F. Encoded NOV8c protein sequence (SEQ ID NO:26).

MRTAPSLRRCVCLLLAAILDLAGYLTVNIEPLPPVAGDAVTLKCNFKTDGRMREIVWYRVTDGGTIKQKIFTDAMFS
TNYSHMENYRKREDLVYQSTVRLPEVRI SDNGPYECHVGIYDRATREKVVLASGNI FLNVMA PPTSIEVVAADT PAFPSR
YQAO NFTLVCIVSGGKPAPMVYFKRDGEPI DAVPLSEPPAASSGPLQDSRPFRSLHRLD DDTKMOKSLSLDAENRGR
PYTERPSRGLTPDPNILLQPTTENI PETVVSREFPRVHSAEPTYFLRHSRT PSSDGTVEVRALLTWT LNPQIDNEALFS
CEVKHPALSMMPMAEVTLVAPKGP KIVMTPSRARVGDTRILVHG FQNEVPPEPMTWTRVGSRLLDGSAEF DGKELVLE
RVPAELNGSMYRCTAQNPLGSTDTHTRLIVFENPNIPRGTEDSNGSIGPTGARLTLVLALT VILELT

The NOV8c amino acid sequence has 401 of 404 amino acid residues (99%) identical to, and 401 of 404 amino acid residues (99%) similar to, a *Macaca fascicularis* 404 amino acid residue protein (ptnr:SPTREMBL-ACC:Q9N0C1) ($E = 1.2e^{-216}$).

NOV8c is expressed in at least the following tissues: Brain, Heart, Thalamus, Lung, Pancreas and Prostate. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV8d

A disclosed NOV8d nucleic acid of 682 nucleotides (also referred to CG50309-04) encoding a novel B7 family-like protein is shown in Table 8G. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 4-6 and ending with a TGA codon at nucleotides 661-663. Putitive untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 8G, and the start and stop codons are in bold letters.

Table 8G. NOV8d Nucleotide Sequence (SEQ ID NO:27)

ACCATGCGGAACCGCCCGAGCCTCCGCGGTGCGCGCCACCGCCTCGGCCAGTGGCCGAGGCAGGAGCGCTGAGTT
TCCCCGCTGGGTCCACAGCGCCGAGCCACCTACTTCTGCGCCACAGCCGACCCCGAGCAGTGACGGCACTGTGGAAG
TACGTGCGCTGCTCACCTGGACCCTCAACCCACAGATCGACAACGAGGCCCTCTCAGCTGCGAGGTCAAGCACCAGCT
CTGTCGATGCCCATGCGGCGAGGTCACGCTGGTTGCCCCCAAAGGACCCAAAATTGTGATGATGCCAGCAGAGCCG
GGTAGGGGACACAGTGAGGATTCTGGTCCATGGGTTTCAGAACGAAGTCTTCCGGAGCCCATGTTACAGTGACGCGGG
TTGGGAGCCGCTCTGGACGGCAGCGCTGAGTTCGACGGGAAGGAGCTGGTGTGGAGCGGGTTCCCGCCGAGCTCAAT

GGCTCCATGTATCGCTGCACCGCCAGAACCCACTGGGCTCCACTGACACGCACACCCGGCTCATCGTGTGTTGAAAACCC
AAATATCCCAAGAGGAACGGAGGACTCTAATGGTTCATTGCCCCCACTGGTGCCCGGCTCACCTTGGTGCTCGCCCTGA
CAGTGATTCTGGAGCTGACGTGATGACAGTGATTCTGGAGCT

The disclosed NOV8d nucleic acid sequence, localized to chromosome 1, has 601 of 653 bases (92%) identical to a *Macaca fascicularis* brain cDNA, clone:QccE-13927 (gb:GENBANK-ID:AB046009|acc:AB046009.1) ($E = 7.7e^{-124}$).

- 5 A disclosed NOV8d polypeptide (SEQ ID NO:28) encoded by SEQ ID NO:27 is 219 amino acid residues and is presented using the one-letter amino acid code in Table 8H. Signal P, Psort and/or Hydropathy results predict that NOV8d does not contain a signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.4500.

Table 8H. Encoded NOV8d protein sequence (SEQ ID NO:28).

MRTAPSLRRCPPPRPVAGGRSASEFPWRVHSAEPTYFLRHSRTPSSDGTVEVRALLTWTLPQIDNEALFSCEVKHPAL
SMPMRAEVTILVAPKGPKI VMMPSRARVGDTRILVHGFQNEVFPPEPMTWTRVGSRLLDGSAEFDGKELVLERVPAELNG
SMYRCTAQNPLGSTDTHTRLIVFENPNI PRGTEDSNGS IAPTGARLTILVLAALTIVILELT

- 10 The NOV8d amino acid sequence has 130 of 132 amino acid residues (98%) identical to, and 131 of 132 amino acid residues (99%) similar to, a *Macaca fascicularis* 404 amino acid residue protein (ptnr:SPTREMBL-ACC:Q9N0C1) ($E = 2.3e^{-70}$).

- NOV8d is expressed in at least the following tissues: Brain, Heart, Thalamus, Lung, Pancreas and Prostate. This information was derived by determining the tissue sources of the
15 sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV8e

- A disclosed NOV8e nucleic acid of 992 nucleotides (also referred to CG50309-05) encoding a novel B7 family-like protein is shown in Table 8I. An open reading frame was
20 identified beginning with an ATG initiation codon at nucleotides 4-6 and ending with a TGA codon at nucleotides 814-816. Putitive untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 8I, and the start and stop codons are in bold letters.

Table 8I. NOV8e Nucleotide Sequence (SEQ ID NO:29)

ACCATGCGAACCGCCCCGAGCCTCCGCGCTGCGTCTGCTGCTGCTCGCCGCGATCCTGGACCTGGCGCGGGCTACCT
GACAGTCAACATTGAGCCTCTCCCCCTGTGGTGGCTGGAGACGCCGTGACTTTGAAGTGTAACCTCAAGACAGATGGGC
GCATGCGGGAGATCGTGTGGTACCGGGTGACGGATGGTGGCACCATCAAGCAAAAGATCTTCACCTTCGACGCCATGTTT
TCCACCAACTACTCACACATGGAGAACTACCGCAAGCGAGAGGACCTGGTGTACCACTGCTGAGGCTGCCCGAGGT
CCGGATCTCAGACAATGGTCCCTATGAGTGCCATGTGGGCATCTACGACCGCGCCACCAGGGAGAAGGTGGTCTGGCAT
CAGGCAACATCTTCTCAACGTCATGGTTGCCCCCAAGGACCCAAAATTGTGATGACGCCAGCAGAGCCCGGGTAGGG
GACACAGTGAGGATTTCTGGTCCATGGGTTTCAGAACGAAGTCTTCCCGAGGCCATGTTACGTTGGACGCGGGTTGGGAG
CCGCCTCTGGACGGCAGCGCTGAGTTCGACGGGAAGGAGCTGGTGCTGGAGCGGGTCCCGCGGAGCTCAATGGCTCCA
TGTATCGCTGCACCGCCCGAACCCTGGGCTCCACGACACGCACACCCGGCTCATCGTGTGTTGAAAACCCAAATATC
CCAAGAGGAACGGAGGACTCTAATGGTTCCATTGGCCCCACTGGTGCCCGGCTCACCTTGGTGCTCGCCCTGACAGTGAT
TCTGGAGCTGACGTGAAGACAGTGATTCTGGAGCTGACGTGACAGTGATTCTGGAGCTGACGTGATGACAGTGATTCTGG

AGCTGACGTGATGACAGTGATTCTGGAGCTGACGTGATGACAGTGATTCTGGAGCTGACGTGATGACAGTGATTCTGGAG
CTGACGTGATGACAGTGATTCTGGAGCTGACG

The disclosed NOV8e nucleic acid sequence, localized to chromosome 1, has 418 of 436 bases (95%) identical to a *Macaca fascicularis* brain cDNA, clone:QccE-13927 (gb:GENBANK-ID:AB046009|acc:AB046009.1) ($E = 5.3e^{-90}$).

5 A disclosed NOV8e polypeptide (SEQ ID NO:30) encoded by SEQ ID NO:29 is 270 amino acid residues and is presented using the one-letter amino acid code in Table 8J. Signal P, Psort and/or Hydropathy results predict that NOV8e contains a signal peptide and is likely to be localized to the lysosome (lumen) with a certainty of 0.8457 and extracellularly with a certainty of 0.6233. The most likely cleavage site for a NOV8e peptide is between amino acids
10 24 and 25, at: ARG-YL.

Although PSORT suggests that NOV8e may be localized in the lysosome (lumen), NOV8e is similar to the B7 family, some members of which are secreted. Therefore it is likely that this NOV8e protein is localized extracellularly.

Table 8J. Encoded NOV8e protein sequence (SEQ ID NO:30).

MRTAPSLRRCVCLLLAAILDLARGYLTVNIEPLPPVAGDAVTLKCNFKTDGRMREIVWYRVTDGGTIKQKIIFTDAMFS
TNYSHMENYRKREDLVYQSTVRLPEVRI SDNGPYECHVGI YDRATREKVVLASGNI FLNVMVAPKGPKIVMTPSRARVGD
TVRILVHGFQNEVFPEPMFTWTRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAPNPLGSTDTHTRLIVFENPNIP
RGTEDSNGSIGPTGARLTLVLALTVILELT

15 The NOV8e amino acid sequence has 153 of 187 amino acid residues (81%) identical to, and 161 of 187 amino acid residues (86%) similar to, a *Macaca fascicularis* 404 amino acid residue protein (ptnr:SPTREMBL-ACC:Q9N0C1) ($E = 7.6e^{-73}$).

NOV8e is expressed in at least the following tissues: Brain, Heart, Thalamus, Lung, Pancreas and Prostate. This information was derived by determining the tissue sources of the
20 sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV8f and NOV8g

Both NOV8c and NOV8e were subjected to "in frame" cloning. The cDNA coding for the mature form of NOV8c (CG50309-03) from residue 25 to 467 was targeted for "in-frame"
25 cloning by PCR. The insert assembly NOVf (also referred to as assembly 169376006) was found to encode an open reading frame between residues 25 and 467 of NOV8c. The NOVf nucleic acid sequence (SEQ ID NO:31) and its corresponding amino acid sequence (SEQ ID NO:32) are shown in Tables 8K and 8L, respectively.

Table 8K. NOV8f Nucleotide Sequence (SEQ ID NO:31)

GGATCCTACCTGACAGTCAACATTGAGCCTCTCCCCCTGTGGTGGCTGGAGACGCCGTGACTTTGAAGTGTAACCTCAAG

ACAGATGGGCGCATGCGGAGATCGTGTGGTACCGGGTGACGGATGGTGGCACCATCAAGCAAAAGATCTTCACCTTCGAC
GCCATGTTCTCCACCACTACTCACACATGGAGAACTACCGCAAGCGAGAGGACCTGGTGTACCACTCCACTGTGAGGCTG
CCCGAGGTCGGATCTCAGACAATGGTCCCTATGAGTGCCATGTGGGCATCTACGACCGCGCCACCAGGAGAAGGTGGTC
CTGGCATCAGGCAACATCTTCTCAACGTCATGGCTCTCCACCTCCATTGAAGTGGTGGCTGCTGACACACAGCCCC
TTCAGCCGCTACCAAGCCAGAACTTCACGCTGGTCTGCATCGTGTCTGGAGGAAAACAGCACCCTATGGTTTATTTCAA
CGAGATGGGGAACCAATCGACGAGTGGCCCTATCAGAGCCACAGCTGCGAGCTCCGGCCCCCTACAGGACAGCAGGCC
TTCCGAGCCTTCTGCACCGTGACCTGGATGACACCAAGATGCAAGTCACTGTCCCTCTGGAGCGCGAGAACCGGGGT
GGGCGACCTTACACGAGCGCCCCCTCCCGTGGCTGACCCAGATCCCAACATCCTCTCCAGCAACACAGAGAACATA
CCAGAGACGGTGGTGGAGCGGTGAGTTTCCCGCTGGGTCCACAGCGCGAGGCCACCTACTTCTGCGCCACAGCGCACC
CCGAGCAGTGACGGCACTGTGGAAGTACGTGCCCTGCTCACTGGACCTCAACCCACAGATCGACAACGAGGCCCTCTTC
AGCTGCGAGGTCAAGCACCCAGCTCTGTGATGCCATGACAGGAGGTGACGCTGGTGGCCCCAAGGACCCAAAT
GTGATGACGCGCCAGCAGAGCCGGGTAGGGGACACAGTGAGGATTCTGGTCCATGGGTTTCAAGAACGAAGTCTTCCGGAG
CCCATGTTACGTTGACGCGGGTTGGGAGCGCCTCTGGAGCGGAGCGCTGAGTTGACGGGAAGGAGCTGGTGTGGAG
CGGGTTCGCCCGAGCTCAATGGTCCATGTATCGTGCACCGCCAGAACCCACTGGGCTCCACCGACACGACCCCGG
CTCATCGTGTGAAAAACCAATATCCCAAGAGGAACGAGGACTCTAATGGTTCATTGGCCCCACTGGTGGCCGGCTC
ACCTTGGTGTGCGCCTGACAGTGATTCTGGAGCTGACGCTCGAG

Table 8L. Encoded NOV8f protein sequence (SEQ ID NO:32).

GSYLTVNI EPLPPVAGDAVTLKCNFKTDGRMREI VWYRVTDGGTI KQKI FTDFAMFSTNYSHMENYRKREDLVYQSTVR
LPEVRI SDNPGYEVHVG IYDRATREKVVLAGNI FLNVMAPPTS IEVVAADTPAPFSRYQAQNF TLVCI VSGGK PAMVY
FKRDGEPI DAVPLSEPPAASSGPLQDSRPFRLHRLDDTKMQKSLSLDAENRGGPRPYTERPSRGLTPDPNILLQPTT
ENIPETVVSREFPRVHSAEPTYFLRHSRTSSDGTVEVRALLTWLNPQIDNEALFSCVVKHPALSMMPQAEVTLVAFK
GPKI VMTPSRARVGDVRI LVHGFQNEVFPPEMFTWTVRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQNPLGST
DTHTRLIVFENPNI PRGTEDSNGSIGPTGARLTLVLALT VILELTLE

The cDNA coding for the mature form of NOV8e (CG50309-05) from residue 25 to 254 was targeted for “in-frame” cloning by PCR. The insert assembly NOV8g (also referred to as assembly 170403925) was found to encode an open reading frame between residues 25 and 254 of the NOV8g target sequence. The NOVg nucleic acid sequence (SEQ ID NO:33) and its corresponding amino acid sequence (SEQ ID NO:34) are shown in Tables 8M and 8N, respectively.

Table 8M. NOV8g Nucleotide Sequence (SEQ ID NO:33)

GGATCCTACCTGACAGTCAACATTGAGCCTCTCCCCCTGTGGTGGCTGGAGACGCCGTGACTTTGAAGTGTAACITCAAG
ACAGATGGGCGCATGCGGAGATCGTGTGGTACCGGGTGACGGATGGTGGCACCATCAAGCAAAAGATCTTCACCTTCGAC
GCCATGTTCTCCACCACTACTCACACATGGAGAACTACCGCAAGCGAGAGGACCTGGTGTACCACTCCACTGTGAGGCTG
CCCGAGGTCGGATCTCAGACAATGGTCCCTATGAGTGCCATGTGGGCATCTACGACCGCGCCACCAGGAGAAGGTGGTC
CTGGCATCAGGCAACATCTTCTCAACGTCATGGTTGCCCCAAAGGACCCAAATTTGATGACGCCAGCAGAGCCCGG
GTAGGGGACACAGTGAGGATTCTGGTCCATGGGTTTCAAGACGAAGTCTTCCCGAGGCCATGTTACGTGGAGCGGGTT
GGGAGCCGCTCTTGGACGGCAGCGCTGAGTTCCAGGGGAAGGAGTGGTGTGGAGCGGGTTCGCCCGAGCTCAATGGC
TCCATGTATCGCTGCACCGCCAGAACCCACTGGGCTCCACCGACACGACACCGGCTCATCGTGTGAAAACCAAAAT
ATCCCAAGAGGAACGGAGGACTCTAATGGTTCATTGGCCCCACTGGTCTCGAG

Table 8N. Encoded NOV8g protein sequence (SEQ ID NO:34).

GSYLTVNI EPLPPVAGDAVTLKCNFKTDGRMREI VWYRVTDGGTI KQKI FTDFAMFSTNYSHMENYRKREDLVYQSTVR
LPEVRI SDNPGYEVHVG IYDRATREKVVLAGNI FLNVMVAPKGPKI VMTPSRARVGDVRI LVHGFQNEVFPPEMFTWT
RVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQNPLGSTDTHTRLIVFENPNI PRGTEDSNGSIGPTGLE

Possible small nucleotide polymorphisms (SNPs) found for NOV8a, NOV8d and NOV8e are listed in Tables 8O, 8P and 8Q, respectively.

Table 8O: SNPs

Variant	Nucleotide Position	Base Change	Amino Acid Position	Base Change
---------	---------------------	-------------	---------------------	-------------

13376224	1039	A > T	346	Met > Leu
13376223	1236	C > T	Silent	N/A

Table 8P: SNPs			
Consensus Position	Depth	Base Change	PAF
381	4	G > A	0.500
429	4	T > C	0.500

Table 8Q: SNPs			
Consensus Position	Depth	Base Change	PAF
436	26	T > C	0.462
438	26	G > C	0.423
450	26	A > C	0.346
451	26	T > C	0.346
472	23	C > G	0.348
475	22	G > C	0.136

NOV8a – NOV8g are very closely homologous as is shown in the amino acid alignment in Table 8R.

Table 8R Amino Acid Alignment of NOV8a – NOV8g

	10	20	30	40	50	60	70
NOV8a	MRTAPSLRRRCVCLLLAAILDLAR	YLTVNIEPLPE	VVAGDAVTLKCN	FKTDGRMREI	VWYRVTDGGTIK		
NOV8b	MRTAPSLRRRCVCLLLAAILDLAR	GYLTVNIEPLPE	VVAGDAVTLKCN	FKTDGRMREI	VWYRVTDGGTIK		
NOV8c	MRTAPSLRRRCVCLLLAAILDLAR	GYLTVNIEPLPE	VVAGDAVTLKCN	FKTDGRMREI	VWYRVTDGGTIK		
NOV8d	-----MRTAPSLRRRCVCLLLAAILDLAR	GYLTVNIEPLPE	VVAGDAVTLKCN	FKTDGRMREI	VWYRVTDGGTIK		
NOV8e	MRTAPSLRRRCVCLLLAAILDLAR	GYLTVNIEPLPE	VVAGDAVTLKCN	FKTDGRMREI	VWYRVTDGGTIK		
NOV8f	-----GSYLTVNIEPLPE	VVAGDAVTLKCN	FKTDGRMREI	VWYRVTDGGTIK			
NOV8g	-----GSYLTVNIEPLPE	VVAGDAVTLKCN	FKTDGRMREI	VWYRVTDGGTIK			
	80	90	100	110	120	130	140
NOV8a	OKIIFTFDAMFS	TNYSHMEN	YRKREDLVYQSTVRLPEVRISDN	GPYECHVG	IYDRATREKVVLASGNI	FI	
NOV8b	OKIIFTFDAMFS	TNYSHMEN	YRKREDLVYQSTVRLPEVRISDN	GPYECHVG	IYDRATREKVVLASGNI	FI	
NOV8c	OKIIFTFDAMFS	TNYSHMEN	YRKREDLVYQSTVRLPEVRISDN	GPYECHVG	IYDRATREKVVLASGNI	FI	
NOV8d	DGTVEVRALLTWTINPOIDN	-----EALFSCSEVKHPALSM	-----M	-----RAEVTI	-----		
NOV8e	OKIIFTFDAMFS	TNYSHMEN	YRKREDLVYQSTVRLPEVRISDN	GPYECHVG	IYDRATREKVVLASGNI	FI	
NOV8f	OKIIFTFDAMFS	TNYSHMEN	YRKREDLVYQSTVRLPEVRISDN	GPYECHVG	IYDRATREKVVLASGNI	FI	
NOV8g	OKIIFTFDAMFS	TNYSHMEN	YRKREDLVYQSTVRLPEVRISDN	GPYECHVG	IYDRATREKVVLASGNI	FI	
	150	160	170	180	190	200	210
NOV8a	NVMAPPTSIEVVAADTPAPFSRYQAQNF	TLVCI	VS	GGK	PAPMVYFKRDGEP	IDAVPLSEPPAASSG	PLQD
NOV8b	NVMAPPTSIEVVAADTPAPFSRYQAQNF	TLVCI	VS	GGK	PAPMVYFKRDGEP	IDAVPLSEPPAASSG	PLQD
NOV8c	NVMAPPTSIEVVAADTPAPFSRYQAQNF	TLVCI	VS	GGK	PAPMVYFKRDGEP	IDAVPLSEPPAASSG	PLQD
NOV8d	-----						
NOV8e	NVM	-----					
NOV8f	NVMAPPTSIEVVAADTPAPFSRYQAQNF	TLVCI	VS	GGK	PAPMVYFKRDGEP	IDAVPLSEPPAASSG	PLQD
NOV8g	NVM	-----					
	220	230	240	250	260	270	280
NOV8a	SRPFRS	LLDDTKMQKSLSL	DAENRGG	RPYTERPS	RGLTPDPNILLQPTTENI	PETVVSREFPRWV	
NOV8b	SRPFRS	LLDDTKMQKSLSL	DAENRGG	RPYTERPS	RGLTPDPNILLQPTTENI	PETVVSREFPRWV	
NOV8c	SRPFRS	LLDDTKMQKSLSL	DAENRGG	RPYTERPS	RGLTPDPNILLQPTTENI	PETVVSREFPRWV	
NOV8d	-----						
NOV8e	-----						
NOV8f	SRPFRS	LLDDTKMQKSLSL	DAENRGG	RPYTERPS	RGLTPDPNILLQPTTENI	PETVVSREFPRWV	

NOV8g

290 300 310 320 330 340 350

NOV8a HSAEPTYFLRHSRT PSSDGTVEVRALLTWTLLNPQIDNEALFSCVEVKHPALSMMPRAEVLTPVAPKGPKI VM

NOV8b HSAEPTYFLRHSRT PSSDGTVEVRALLTWTLLNPQIDNEALFSCVEVKHPALSMMPRAEVLTPVAPKGPKI VM

NOV8c HSAEPTYFLRHSRT PSSDGTVEVRALLTWTLLNPQIDNEALFSCVEVKHPALSMMPQAEVTLVAPKGPKI VM

NOV8d -----VAPKGPKI VM

NOV8e -----VAPKGPKI VM

NOV8f HSAEPTYFLRHSRT PSSDGTVEVRALLTWTLLNPQIDNEALFSCVEVKHPALSMMPQAEVTLVAPKGPKI VM

NOV8g -----VAPKGPKI VM

360 370 380 390 400 410 420

NOV8a TPSSRARVGDTVRILVHGFQNEVFFPEPMFTWTRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQN

NOV8b TPSSRARVGDTVRILVHGFQNEVFFPEPMFTWTRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQN

NOV8c TPSSRARVGDTVRILVHGFQNEVFFPEPMFTWTRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQN

NOV8d MPSSRARVGDTVRILVHGFQNEVFFPEPMFTWTRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQN

NOV8e TPSSRARVGDTVRILVHGFQNEVFFPEPMFTWTRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQN

NOV8f TPSSRARVGDTVRILVHGFQNEVFFPEPMFTWTRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQN

NOV8g TPSSRARVGDTVRILVHGFQNEVFFPEPMFTWTRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRCTAQN

430 440 450 460 470

NOV8a LGSTDTHTRLIVFENENIPRGTEDSNNGSIGPTGARLTLVLALTIVILELT

NOV8b LGSTDTHTRLIVFENENIPRGTEDSNNGSIGPTGARLTLVLALTIVILELT

NOV8c LGSTDTHTRLIVFENENIPRGTEDSNNGSIGPTGARLTLVLALTIVILELT

NOV8d LGSTDTHTRLIVFENENIPRGTEDSNNGSIGAPTGARLTLVLALTIVILELT

NOV8e LGSTDTHTRLIVFENENIPRGTEDSNNGSIGPTGARLTLVLALTIVILELT

NOV8f LGSTDTHTRLIVFENENIPRGTEDSNNGSIGPTGARLTLVLALTIVILELT

NOV8g LGSTDTHTRLIVFENENIPRGTEDSNNGSIGPTG-----LE---

Homologies to any of the above NOV8 proteins will be shared by the other NOV8 proteins insofar as they are homologous to each other as shown above. Any reference to NOV8 is assumed to refer to both of the NOV8 proteins in general, unless otherwise noted.

5 The disclosed NOV8a polypeptide also has homology to the amino acid sequences shown in the BLASTP data listed in Table 8S.

Table 8S. BLAST results for NOV8a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14249630 ref NP_116269.1 (NM_032880)	hypothetical protein MGC15730 [Homo sapiens]	467	462/467 (98%)	463/467 (98%)	0.0
gi 9280074 dbj BAB01591.1 (AB046009)	unnamed protein product [Macaca fascicularis]	404	389/404 (96%)	391/404 (96%)	0.0
gi 7263981 emb CAB81618.1 (AL050342)	dJ655K7.1 (novel protein) [Homo sapiens]	80	80/80 (100%)	80/80 (100%)	3e-42
gi 11359852 pir T42633	connectin/titin [Gallus gallus]	4162	80/366 (21%)	146/366 (39%)	2e-05
gi 14575679 gb AAK68690.1 AF156100_1 (AF156100)	hemacentin [Homo sapiens]	5636	88/357 (24%)	150/357 (41%)	9e-05

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 8T.

Table 8T. ClustalW Analysis of NOV8a

- 2) [gi|14249630|ref|NP_116269.1](#) (NM_032880) hypothetical protein MGC15730 [Homo sapiens] (SEQ ID

NO:89)

- 3) gi|9280074|dbj|BAB01591.1| (AB046009) unnamed protein product [Macaca fascicularis] (SEQ ID NO:90)
4) gi|7263981|cemb|CAB81618.1| (AL050342) dJ655K7.1 (novel protein) [Homo sapiens] (SEQ ID NO:91)
5) gi|11359852|pir|T42633 connectin/titin [Gallus gallus](SEQ ID NO:92)
6) gi|14575679|gb|AAK68690.1|AF156100.1 (AF156100) hemicentin [Homo sapiens] (SEQ ID NO:93)

```

      10      20      30      40      50      60      70
...|...|...|...|...|...|...|...|...|...|...|...|...|
NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| -MTTKAPTFTQPLQSVVALEGSAAATFEAHISGFPVPEVSWYRDGQVLSAATLPGVQISFSDGRAKLVI PS
gi|14575679| MISWEVHTVFLFALLYSSLAQDASPQSEIRAEFPPEGASTLAFVFDVTGSMYDDLQVIEGASKILETS

      80      90     100     110     120     130     140
...|...|...|...|...|...|...|...|...|...|...|...|
NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| VTEANSGRYTIQATNGSGQATSTAELLVTAGTAPPNFSQRLQSMTARQGSQ-----VRLDVVRTG
gi|14575679| LKRPKRPLPNFALVP----FHDPEIGPVTITTDPKKFQYELRELYVQGGDCPEMSIGAIKIALEISLPG

      150     160     170     180     190     200     210
...|...|...|...|...|...|...|...|...|...|...|...|
NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| IPTPVVKFYRDGVEIQSSPDFQILQEGDLYSLIIAEAYPEDSGTYSVNATNNVGRATSTAELLIQGEIEEA
gi|14575679| SPIYVFTDARSKDYRLTHEVLQLIQKQSQVVFVLTGDCDDRTHIGYKYVEEIASTSSGQVFHLDKKQVNV

      220     230     240     250     260     270     280
...|...|...|...|...|...|...|...|...|...|...|...|
NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| VPAKTKTKTIVSTAQISQTRQARIEKKIETHFDARSLTSVEMVIEGAAAQQLPHKAPPR-----
gi|14575679| EVLKWVEEAVQASKVHLLSTDHLEQAVNTWRIFPDPSLKEVTVSLSGPSPMIEIRNPLGKLIKKGFGLHE

      290     300     310     320     330     340     350
...|...|...|...|...|...|...|...|...|...|...|...|
NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| -----MPPRPTSKSPTPPVITAKAQMARQQSPSPVRQSPSPVRHVRAPT
gi|14575679| LLNIHNSAKVVNVKEPEAGMWTVKTSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSRPVQGIPTIV

      360     370     380     390     400     410     420
...|...|...|...|...|...|...|...|...|...|...|...|
NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| PSPVRSVSPAGRISTSPIRPVKSPS-----PIR-----KAQVVTPGAEVLPWR-----QEGYSAT
gi|14575679| LLNTSGISTPARIDLLELLSISGSSLKTI PVKYYPHRKPYGIWNISDFVPPNEAFFLKVTGYDKDDYLFQ

      430     440     450     460     470     480     490
...|...|...|...|...|...|...|...|...|...|...|...|
NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| AEAQMKETRVSTSAEIRTEERWEGRYGLQEQVTISG-----AAAGEVAAGAKEVRKEPEKTPVPT
gi|14575679| RVSSVSFSSIVPDAPKVTMPEKTPGYLQGGQIPCSVDSLLPFTLSFVRNGVTLGVDQYLKESASVSLDI

      500     510     520     530     540     550     560
...|...|...|...|...|...|...|...|...|...|...|...|
NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| VIIATDKAKEQERISTAR-----EEISARHEQVHVS-HEQIEAGKRAEAVATVVAADV-----
```


[illegible]

NOV8a

NOV8a

NOV8a

NOV8a

NOV8a

NOV8a

NOV8a

NOV8a

NOV8a

gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|---SQ-----QE---MMYQTQVTAYVQEPK--VAEVAPPISYGDGDFDKEYEKEQ
gi|14575679|TQLISFFSPRHTFLPSGSMKITETRTSDSGMYLCVATNIAGNVTQAVKLVNHVPPKIQRGPKHLKVQVQG

1200 1210 1220 1230 1240 1250 1260
...|...|...|...|...|...|...|...|...|...|...|...|...|

NOV8a
gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|ALIRKKMAKDTVMVR-----TFVEDEEFHISSFEERLIKEIELR---I I KTTLDELLEEDGEEEMMI
gi|14575679|RVDIPCNAQGTPLPVITWSKGGSTMLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEITL

1270 1280 1290 1300 1310 1320 1330
...|...|...|...|...|...|...|...|...|...|...|...|...|

NOV8a
gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|DISESEAIG-AGFDLRLKNYRTFEGTGVTFHCKTTGYPLPKIAWYKDGKRI RHGERYHMEVLQDGSASLR
gi|14575679|HVQEPPTVEDLEPPYNTTFQERVANQRIEFPCKPAKGT PKPTIKWLHNGREL-TGREPGISILEEG-TLLV

1340 1350 1360 1370 1380 1390 1400
...|...|...|...|...|...|...|...|...|...|...|...|...|

NOV8a
gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|LPVVLPEDEGIYTVFASNMGK--NAICSAKLYVEPVAPTATPG-----YMPGPEVMR
gi|14575679|IASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKDKEQVSNVSVLLNQLTNLFCVEGTPSPIIMW

1410 1420 1430 1440 1450 1460 1470
...|...|...|...|...|...|...|...|...|...|...|...|...|

NOV8a
gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|RYRSISPRSPSRSPARSSPSCSPARRLDETDEG-----QLERLYKPVFVLK--PTSVK
gi|14575679|YKDNVQVTESSITQTVNNGKILKLFRTPEDAGRYSCKAINIAGTSQKYFNIDVLVPPTIIGTNFPKEVS

1480 1490 1500 1510 1520 1530 1540
...|...|...|...|...|...|...|...|...|...|...|...|...|

NOV8a
gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|CSQGQTARFDLKVVGPRMPETYWFHNGQQVNDYTHKIVIKEDGTQSLIIVPAMPEDSGEWAVIAQNRAG
gi|14575679|VVLNRDVALECVKGT PFPDIHWFKDGKPLFLGDPNVELL--DRGQVLHLKNARRNDKGRYQCTVSNAG

1550 1560 1570 1580 1590 1600 1610
...|...|...|...|...|...|...|...|...|...|...|...|...|

NOV8a
gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|KASVSVTL SVEAKEDLVRPRFVERLRNVSVKEGSR LHMVAVKATGNPNPDIVWLKNSDIIVPHKYPRIE
gi|14575679|KQAKDIKLTIIYIPPSIKGGNVTT--DISVLINSLIKLECKTRGLPMPAITWYKQGQPI MSSQALYIDK

1620 1630 1640 1650 1660 1670 1680
...|...|...|...|...|...|...|...|...|...|...|...|...|

NOV8a
gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|GTKGAAALNIESTARQDAAWYTATAINKAGRDTRCKVN-----VEVEHAEPEPERRLIIPKGYKAKE
gi|14575679|GQY---LHIPRAQVSDSATYTCHVANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECN

1690 1700 1710 1720 1730 1740 1750
...|...|...|...|...|...|...|...|...|...|...|...|...|

NOV8a
gi|14249630|-----
gi|9280074|-----
gi|7263981|-----
gi|11359852|IAAPELEPLHLRYG-----Q-----EQWEEGD-----LYDKEKQKPPF
gi|14575679|AGNPSPIITWLKDGVPVKANDNFRIEAGGKKLEIMSAQEIDRGQYICVATSVAGEKEIKYEVDVLVPPAI

1760 1770 1780 1790 1800 1810 1820
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|
 gi|9280074|
 gi|7263981|
 gi|11359852| KKKLTSRLKQFGPAHFECRLTPIGDPTMVVEWLHDGKPLEAANRLRMINEFGYCSLDYGVAYSRDSGVI
 gi|14575679| EGGDETSYFIVMVNLELDCHVTGSPPTIMWLKDGQLIDERDGFKIL--LNGRKLVIQAQVSNLTGLY

1830 1840 1850 1860 1870 1880 1890
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|
 gi|9280074|
 gi|7263981|
 gi|11359852| TCRATN-----KYG--TDHT-SATLIVKDE-----KSLVEESQLPEGRGM
 gi|14575679| RCMAANTAGDHKKEFEVTVHVPPTIKSSGLSERVVVKYPVALQCIANGIPNPSITWLKDDQPVNTAQGN

1900 1910 1920 1930 1940 1950 1960
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|
 gi|9280074|
 gi|7263981|
 gi|11359852| ORIEELER-----MAHEGALPAVAVDQK-EKQKPELVLPVPEPARVLEG-----ETARFRCR
 gi|14575679| LKIQSSGRVLQIAKTLLDAGRYTCVATNAAGETQQHIQLHVHEPPSLEDAGKMLNETVLVSNPVQLECK

1970 1980 1990 2000 2010 2020 2030
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|
 gi|9280074|
 gi|7263981|
 gi|11359852| VTGYPLPKVNWYLSQLIRKSKRFLRYDGIHYLDIVDCKSYDTGEVKVTAENPEGFI EHKVKLEIQORE
 gi|14575679| AAGNPVPVITWYKDNCLLSGSTSMTF-LNRGQIIDIESAQISDAGIYKCVAINSAGATELFYSLQVHVAP

2040 2050 2060 2070 2080 2090 2100
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|
 gi|9280074|
 gi|7263981|
 gi|11359852| DFRSVLRRAPRHEPVVTE-----PG-KLLFEVQKIDKPAEATTKEVVKLKRAERITHEKLSESEEL
 gi|14575679| SISGSNNMVAVVVNNPVRLECEARGIPAPSLTWLKDGSFVSSFNGLQVLGGGRILALSTQISDTGRYT

2110 2120 2130 2140 2150 2160 2170
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|
 gi|9280074|
 gi|7263981|
 gi|11359852| RSKFKRRTEEGYEAITAVELKSRKKDES YEMLKKTKEELLHWTKEIPE--E--E-----KKALPP
 gi|14575679| CVAVNAAGEKQRDIDL RVYVPPNIMGEEQNVSVLISQAVELLQSDAIPPTLTWLKDGHPPLKKPGLSI

2180 2190 2200 2210 2220 2230 2240
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|
 gi|9280074|
 gi|7263981|
 gi|11359852| EGKITIPTFKPEKVELSP--SMEAPKIFER-----IQSQTVAQGTDAHFRVRVV
 gi|14575679| SENRSVLKIEDAQVQDTGRYTCEATNVAGKTEKKYNNVNIWVPPNIGGSDELTLQTVIEGNLISLLCESS

2250 2260 2270 2280 2290 2300 2310
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|
 gi|9280074|
 gi|7263981|
 gi|11359852| GKPDPECQWFRNGVQIERTDRIYWWPEDNVCELVIRDVTADDSASIMVKAVNIAGETSSHAFLLVQAKQ
 gi|14575679| GIPPNLIWKKKGSPVL-TDSMGRVRILSGGRQLQISIAEKSDAALYSCVASNVAGTAKKEYNLQYIIRP

2320 2330 2340 2350 2360 2370 2380
|....|....|....|....|....|....|....|....|....|....|....|
 NOV8a
 gi|14249630|

gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

LISFIQNLQDVVAKERDSMATFECETS-EPFIKVKWFKNGIEIHSGEKYRMHSDRKAHFLSVLAVEMSDA
TITNSGSHPTETIIVTRGKSISLECEVQGIPTPTVTWMKDGHPILIKAGVEILDEG--HILQLKNIHVSDT

NOV8a

gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

2390 2400 2410 2420 2430 2440 2450
....|....|....|....|....|....|....|

DDYSCALVEDE--SVKTTAKLIVEGAVVEFIKELEDVEVPESFTGELECEVSPEDIEG-KWYHGDVELS
GRYVCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVVEKNSVSLTCEASGILPSTTWFKDGPVVS

NOV8a

gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

2460 2470 2480 2490 2500 2510 2520
....|....|....|....|....|....|....|

SNHKYVLASRRGRRLTIKDVNKDDQGEYSFVVD---GKRTHCKLKMKPRPMTILQG-LTDQKVCEDGI
LSNSVRIL--SGGRMLRLMQTTMEDAQYTCVVRNAAGEERKIFGLSVLVPPHIVGENTLEDVKVKEKQS

NOV8a

gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

2530 2540 2550 2560 2570 2580 2590
....|....|....|....|....|....|....|

VQLEVKVSVENVEG-VWMKDGHEIQSSDRIHIVLDKQSHMLLIEDATQEDSGTYSFSIPGLELSTTGQVT
VTLTCEVTGNPVPETIWHKDGQPLQDEAHHIISGGRFLQITNVQVPHTGRYTCLASSPAGHKRSRFSLN

NOV8a

gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

2600 2610 2620 2630 2640 2650 2660
....|....|....|....|....|....|....|

VYSVEIIVP-----LKDVHVVEGTKAILECKVSAPDVTSSKWYLNHDQIKPDERVQAVCKGKQRLVIT
VFSPTIAGVSGDGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTPLESNRIRILPGGRTLQILNA

NOV8a

gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

2670 2680 2690 2700 2710 2720 2730
....|....|....|....|....|....|....|

RTHAS--DEGHYKLMVGKVETSCNVTVEEIEIIRG-----LHDITCTETQNVSPFEVELSHSG-IDV
QEDNAGRYSCVATNEAGEMIKHYEVKVYIPPIINKGDLWGPGLSPKEVKIKVNNLTLECEAYAIPASAL

NOV8a

gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

2740 2750 2760 2770 2780 2790 2800
....|....|....|....|....|....|....|

IWHFKGQEIKAGPKYKIEARGKIYKLTVVKMMK-----DDEGEYVFYAGGKKTSGKLIIVAGG
SWYKDGQPLKSDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELDFDVNIQVPPSFQKLWEIGN

NOV8a

gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

2810 2820 2830 2840 2850 2860 2870
....|....|....|....|....|....|....|

AIS---KPLADLTVAESQRAVFECEVANPESRGQWLKNGKPLPMTDQYRAETDGVKRRLLNIPAAKMDDMG
MLDTGRNGEAKDVIINNPIISLYCETNAAPPPTLTWYKDGHPILTSSDKVLILPGGRV--LQIPRAKVEDAG

NOV8a

gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

2880 2890 2900 2910 2920 2930 2940
....|....|....|....|....|....|....|

MRTAFSLR
MRTAFSLR
MRTAFSLR
EYSYEIASSKTSAKLHVEVVKIK
RYTCVAVNEAGEDSLQYDVRVLVPPIIIEGANGDLPEEVTVLVNKSALIECLSSGSEAPRNSWQKDGQPLL

NOV8a
 gi|14249630|
 qi|9280074|

QDSRPFRSLLL-DEDDTKMQKSLSLLDABNRGRGPYTERESRGLTPDPNNILQPTTENIPETVWSREFPR
 QDSRPFRSLLHRDLDTKMQKSLSLLDABNRGRGPYTERESRGLTPDPNNILQPTTENIPETVWSREFPR

gi|7263981|
gi|11359852|
gi|14575679|

3580 3590 3600 3610 3620 3630 3640

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

3650 3660 3670 3680 3690 3700 3710

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

3720 3730 3740 3750 3760 3770 3780

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

3790 3800 3810 3820 3830 3840 3850

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

3860 3870 3880 3890 3900 3910 3920

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

3930 3940 3950 3960 3970 3980 3990

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

4000 4010 4020 4030 4040 4050 4060

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

4070 4080 4090 4100 4110 4120 4130

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679|

4140 4150 4160 4170 4180 4190 4200

74

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
|.....|.....|.....|.....|.....|.....|.....|.....|.....|
 -----AVIVPLRDVAVTSEGG-----SARFCRVTG-----TDLKVS
 PITLSCEADGLPPPDITWHKDGRAIVESIRQRVLSSGSLQIAFVOPGDAGHYTCMAANVAGSSSTSKLT

4210 4220 4230 4240 4250 4260 4270
 . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . |

 WYSDKREIKPSRFFRMTOQFEDTYQLEIAEAYPE---DEGTYTFVASNSVSGQVTSTAILKLEAP-----
 VHVPPRIIRSTEGHYTVNENSOAILPCVADGIPTPAINWKKNVDLLANLLGKYTABPYGLIILENVVLDS

----- EKIMYEKLEEEIEMEVKVAPILRRRRLPLEVAVNHHAKFTCEVETTPNVKFQWYKAGREIYDGD
GFTYCVANNAACEDTHTVSLTVHVLPTFTELPGDVSLNKGEOLRLSCKATGIPLPKLTWTENNNIIPAA--

4350 4360 4370 4380 4390 4400 4410
|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....

 KYSIRSSNYLSTLEIPRPQVWDCEGYSCASQNHGVSSTAFLTVTPEPRFIKKLSDSSRLVKQHDSTRYE
 --HFDSVNGHSELVIERVSKEDSGTYVCTAENSVGFKATGFVYVKEPPVFKGDYPSNWIETPIGCGNAIIN

4420 4430 4440 4450 4460 4470 4480
|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....
 CKVGGSP EIKVTYWKETEIH PSEKYSMSFVDSDAVLEMHNL SVEDSGDYSC EAQNPAGSASTSTSLKV K
 CEVKGDPTPTI QNRRKGV D I EISHRI R ---LNGSLAI YGT V NEDAGD YTCVATNEAGVVERSM SLT L O

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
 APPAFTKKPHPVQTLKGSVDHLECELGQTPEPFQISWYDKDKEIRSSKKYKVMSENYLASIHILNVDITADV
 SPPIITLEPVEVTVINAGGKIILNCOATGEPOPTITWSROGHISIWDNRVNVLNNNSLY---LADAKEDT

4560 4570 4580 4590 4600 4610 4620
|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....

 GEYHCKAVNDVGS DSCIGSVTLRAPPTFVKKLS--DVTVVVVGETIELQAAVEG-----AQPIS
 SEFECAVRNLMGSLVLRVPVIIVOHGGFSOWSAWRACSVTCCKGIOKRSRLCNPLPANGGKPCQGSDFE

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
4630 4640 4650 4660 4670 4680 4690
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----
V/LWLKDKGEEI IRESENLSWISYSENVASLKGICNAEPTN-----AG--K--YICQIKNDAG--
MRNCONKPCPDVGSGSWSEWSLWEECTRSCGRGNOTRTTRTCNNPSVOHGGRPCEGNAVEITMCNIRPCPVHG

100-3634-323702

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

 4770 4780 4790 4800 4810 4820 4830

KDGNELSTDHKYKISFFNKVSGLKILNAGLE - - - - DSGEYTFEVKNVSGKSSCTAS - - - LQVS - - -
SCGGGARQRTRGCDPVPFYGGRCCEGDVQSDFCNSDPCTHGNWSPWSGWGTCSRTCNGGQMRRYRTC

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....

DRIMPPSFTRKLKETYGQLGSSAVLECKVYGSPPILVSWFHDGQEITSGDKYQATLTDN-----TCS
 DNPPPSNGGRACGGPDSQIQRCNTDMCPVDGSGWSGHSWGSQCSASCGGGEKTRKRLCDHPVPVKGGRPCP

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
 LKVNGLQES-----D---MGTYSCTATNVAGSDECSAFLSVREPP-----
 GDTTQTTRCNVQACPGGPQRARGSVIGNINDVEFGIAFLNATITDSPNSDTRIIRAKITNVPRSLGSAMR

.....4980.....4990.....5000.....5010.....5020.....5030.....5040

-----SFVKKPEPFNVLSGENITFTSIVK-----
KIVSILNPIYWTTAKEIGEAVNGFTLTNAVFKRETQVEFATGEILQMSHIARGLDSDGSLLLDIVVSGYV

```

      5050      5060      5070      5080      5090      5100      5110
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
-----|-----|-----|-----|-----|-----|-----|-----|-----|
-----|-----|-----|-----|-----|-----|-----|-----|-----|
-----|-----|-----|-----|-----|-----|-----|-----|-----|
-G--SPP--LEVKWFRGSIELAPGHKCN-----ITLQDSVA
LQLQSPAENVTKDYTEDYIQTGPGQLAYASTRLFTIDGISIPYTNWHTVFYDQAQRMPFLVETLHASSV

```

.....5120.....5130.....5140.....5150.....5160.....5170.....5180

ELELFDVQPLQSGDYTCQVSNEAGKISCTTHLFVKPEPAKFVMKVNDLSVEK---GKNLILECTYGTGPPI
ESDYNQIEETLGFKIHASISKGDRSNQCPSGFTLDSVGPFCAEDEDECAAGNPCSHSCHNAMGYTYCSCP

5190 5200 5210 5220 5230 5240 5250

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

SVTWKKNQVILKHSEKSIITTTETSAILEIPNSKLEDQGGYSCHIENDSGQDNCHGAITILEPPYFVVTPL
GLTIAADGRTCQDIDECALGRHTCHAGQDCDNT----IGSYRCVVRCGSGFRRTSDGLSCQDINECQESS

5260 5270 5280 5290 5300 5310 5320

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

EPVQVTVGDSASLQCQVAGTPEMIVSWYKGDTKLRG-TATVKMHFKN-----QVATLVF

PCHQRCFNAIGSFHCGCEPGYQLKGRKCMDVNECRQNVCRPDQHCNTRGGYKCIDLCPNGMTKAENGTC

76

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| SQVDSDDSGEYICKVENTVGEATSSSLTQERKLPPSFTRKLRDVHETVGLP
gi|14575679| IDIDECKDGTQCRYNQICENTRSSYRCVCPRGYRSQGVGRPCMDIDECEQVPKPCAHQCSNTPGSFKCI

5400 5410 5420 5430 5440 5450 5460
.....|.....|.....|.....|.....|.....|.....|

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| -----VTFDCGIAGSEPIEVSFKNVVRKEDYNVHTSFIDN-----VAILQILK
gi|14575679| CPPGQHLLGDGKSCAGLERLPNYGTQYSSYNLARFSPVRNNYQPPQHYRQYSHLYSSYSEYRNSRTSLSR

5470 5480 5490 5500 5510 5520 5530
.....|.....|.....|.....|.....|.....|.....|

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| TDKS-----LMGQVTCASN-----AIG
gi|14575679| TRRTIRKTCPEGSEASHDTCVDIDECENTDACQHECKNTFGSYQCI CPPGYQLTHNGKTCQDIDECLEQN

5540 5550 5560 5570 5580 5590 5600
.....|.....|.....|.....|.....|.....|.....|

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| TASSSGKLVLTGKTPPFDDTPI TP--VDGIIG-----ESADFECHISG-----TQPIRVTWAKD
gi|14575679| VHCGFNRMCFNMRGSYQCIDTPCPPNYQRDPVSGFCLKNCPPNDLECALSPYALEYKLVSLPFGIATNQD

5610 5620 5630 5640 5650 5660 5670
.....|.....|.....|.....|.....|.....|.....|

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852| NQ-----
gi|14575679| LIRLVAYTQDGMHPRTTFLMVDEEQTVPFALRDENLKGVVYTRPLREAETYRMRVRASSYSANGTIEY

5680
.....|.....|.....|

NOV8a
gi|14249630|
gi|9280074|
gi|7263981|
gi|11359852|
gi|14575679| QTTFIVYIAVSAYPY

Tables 8U and 8V list the domain description from DOMAIN analysis results against NOV8a. This indicates that the NOV8a sequence has properties similar to those of other proteins known to contain these domains.

Table 8U Domain Analysis of NOV8a

gnl|Smart|smart00408, IGC2, Immunoglobulin C-2 Type (SEQ ID NO:94)
Length = 63 residues, 100.0% aligned
Score = 40.4 bits (93), Expect = 2e-04

NOV8a: 352 RVGDTVRIILVHGFQNEVFPEPMFTWTWRVGSRLLDGSAEFDGKELVLERVPAELNGSMYRC 411
|++| + + | | | + | | + | | ++ | | + | + | |
00408: 1 LEGESVTLTCPASGD---PVPNITWLKDGKPLPESRVVASGSLTITKNVSLEDSG-LYTC 56

NOV8a: 412 TAQNPLG 418
|+| +|
00408: 57 VARNSVG 63

gnl|Smart|smart00409, IG, Immunoglobulin (SEQ ID NO:95)
Length = 86 residues, 94.2% aligned
Score = 38.9 bits (89), Expect = 7e-04

B7 molecules play crucial roles in T-cell activation making them plausible targets for cancer, AIDS, and inflammation therapies. The NOV8 proteins described here is known to be expressed in Brain, Heart, Thalamus, Lung, Pancreas, and Prostate tissue, which may indicate a roles in brain and CNS disorders, endocrine, inflammation and autoimmune disorders, pancreatic disorders, and cancers including lung, pancreas, brain, and prostate.

Despite the fact that many tumors express MHC class I molecules presenting "foreign" peptide antigens, a vigorous tumor-destructing immune response is seldom detected. A possible explanation is that tumors cannot provide adequate costimulatory signals as provided by professional antigen presenting cells. CD28, upon interacting with B7, triggers costimulatory signals critical for the T-cell response. Transfection of tumor cells with B7 augments the immunogenicity of the tumor so that an anti-tumor immune response can be amplified. When B7-CD28 costimulation is provided cytotoxic T-lymphocyte (CTL) specific for otherwise silent epitopes can be activated. Therefore, unresponsiveness of T cells to many tumor antigens should be considered as ignorance rather than tolerance. Immunological ignorance may thus contribute to the failure of the immune system to respond against the tumor antigens (Melero et al., Costimulation, tolerance and ignorance of cytolytic T lymphocytes in immune responses to tumor antigens. Life Sci 60(23):2035-41, 1997).

To generate a CTL response to cancer cells requires tumour-specific antigens appropriately processed and displayed by the MHC proteins; T-lymphocytes with receptors of appropriate specificity to recognise these; and initial antigen presentation to the immune system in an immunogenic context. In vitro, autologous tumour-specific CTL have been raised against a number of tumours, thus at least some patients have a suitable combination of antigen and receptor. Vaccination with antigen, or with DNA or viral vectors encoding the antigen, leading to the presentation of identified antigens in an immunogenic context, can activate T-cells which provide protection from tumour in animal models. An alternative approach uses gene transfer to T-cells, causing them to express novel receptors which direct

their cytotoxic activity towards the tumour. Recent advances in understanding the requirements for T-cell activation suggest that failure to efficiently present antigen in an immunogenic context may explain the apparent lack of tumour-specific CTL activation in vivo. In mice, expression of the costimulatory molecule B7-1 on tumour cells, following gene transfer, allows the modified tumour cells to act as antigen-presenting cells, inducing protective and therapeutic CTL responses in some cases (Searle and Young, Immunotherapy II: Antigens, receptors and costimulation. Cancer Metastasis Rev 15(3):329-49, 1996).

Systemic lupus erythematosus antigen-presenting cells fail to upregulate the expression of B7-1 (CD80) in response to interferon gamma; defective expression of B7-1 is responsible for the decreased response of lupus cells to recall antigens (Tsokos, Lymphocytes, cytokines, inflammation, and immune trafficking. Curr Opin Rheumatol 8(5):395-402, 1996).

There is considerable evidence to support an important role for co-stimulatory molecules in regulating the proliferation and activation of T cells in the immune response. Of particular relevance is the interaction between CD28 on T cells and B7 expressed on the surface of antigen presenting cells (APCs). CTLA-4, another molecule present on activated T cells may downregulate T cell activity, but its role remains uncertain. CTLA4-Ig, a fusion protein consisting of the extracellular domain of CTLA4 and the Fc portion of human immunoglobulin G1 (IgG1), has been useful for studying the role of CD28/B7 interactions in immune responses. A number of studies have shown that CTLA4-Ig can switch off T cell activation. Anti B7-2 treatment has similar effects suggesting that interaction of B7-2 with CD28 is important in the development of a Th-2 type inflammatory response in mice. Recent observations have been of relevance to human allergic disease. In vitro studies have shown that CTLA4-Ig or anti-B7-2 antibody can inhibit allergen-induced proliferation and cytokine production by peripheral blood mononuclear cells from atopic subjects. The role of co-stimulation has been studied in a human bronchial explant model of asthma. CTLA4-Ig fusion protein effectively blocked allergen-induced production of IL-5 and IL-13 in bronchial explants from atopic asthmatics. These studies confirm the requirement for interaction between co-stimulatory molecules in cytokine production and allergic inflammation, and point to the CD28-B7 pathway as being important to the allergen-induced inflammation in asthma. Studies of organ transplantation in primates suggest that CTLA4-Ig is extremely effective in preventing organ rejection (Djukanovic, The role of co-stimulation in airway inflammation. Clin Exp Allergy. 230 Suppl 1:46-50, 2000).

Therefore an immune-based gene therapy strategy was selected in which the tumors were transfected with the gene for an alloantigen, human leukocyte antigen (HLA)-B7, a class

I major histocompatibility complex (MHC). This would restore an antigen presentation mechanism in the tumor to induce an antitumor response. Significant advances have been made in the field of gene therapy for cancer. Alloantigen gene therapy has had efficacy in the treatment of cancer and can induce tumor responses in head and neck tumors. Alloantigen gene therapy has significant potential as an adjunctive treatment of head and neck cancer.

The contribution of B7 co-stimulation to CD4+ responses depends upon the activation history of the T-cell and the strength of the T-cell antigen receptor signal. B7 co-stimulation contributes to interleukin (IL)-2 production by both naive and previously activated CD4+ T cells. B7 co-stimulation is most critical for the differentiation of naive CD4+ T cells to IL-4 producers, but predominately influences IL-2 production by previously activated CD4+ cells. B7 co-stimulation is important in development of cytotoxic T cells through both effects on T-helper cells and by direct co-stimulation of CD8+ cells (McAdam et al., The role of B7 co-stimulation in activation and differentiation of CD4+ and CD8+ T cells. *Immunol Rev* 1998 Oct;165:231-47, 1998).

The current model of T cell activation requires two signals. The first signal is specific, requiring T cell receptor recognition and binding to MHC/Antigen presented by an antigen-presenting cell. The second signal is nonspecific, resulting from the binding of B7 ligand on the antigen-presenting cell with its receptor, CD28, on the T cell. If both signals are provided, the T cell will proliferate and secrete cytokines. Recently, it has been shown that CTLA4, another receptor for B7 that is upregulated following T cell after activation, can deliver an inhibitory signal, downregulating T cell proliferation. The B7 family of ligands has two family members, B7-1 and B7-2. They both bind to CD28 and CTLA4, but they differ in their binding affinity, structure, and temporal expression. Considerable research has been done on the CD28/B7 costimulatory pathway. Different ways of manipulating this pathway could provide insights into the mechanism and treatment of opposing pathological states. Blocking the CD28/B7 pathway could result in immunosuppression, with implications for the treatment of autoimmune diseases, organ transplantation, and graft vs. host disease. Activating the CD28/B7 pathway could be useful for including the immune system to recognize and eliminate tumors that evade the immune system. Finally, the CD28/B7 pathway could be involved with maintaining immune tolerance, as recent studies suggest the preferential binding of the B7-CTLA4 pathway results in the down-regulation of the responding T cells. Thus, the B7/CD28/CTLA4 pathway has the ability to both positively and negatively regulate immune responses (Greenfield et al., CD28/B7 costimulation: a review. *Crit Rev Immunol* 1998;18(5):389-418, 1998).

The initiation and progression of autoimmune diseases, such as insulin-dependent diabetes mellitus (IDDM), are complex processes that depend on autoantigen exposure, genetic susceptibility, and secondary events that promote autoaggression. T-cell costimulation, largely mediated by CD28/B7 interactions, is a major regulatory pathway in the activation and differentiation of T-cells that cause IDDM in murine models (Herold et al., CD28/B7 regulation of autoimmune diabetes. Immunol Res. 16(1):71-84, 1997; Toes et al., CD40-CD40Ligand interactions and their role in cytotoxic T lymphocyte priming and anti-tumor immunity. Semin Immunol. 10(6):443-8, 1998).

The above defined information for NOV8 suggests that NOV8 may function as a member of a B7 protein family. Therefore, the NOV8 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the NOV8 compositions of the present invention will have efficacy for treatment of patients suffering from: brain disorders including epilepsy, eating disorders, schizophrenia, ADD, cancer; heart disease, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders, psoriasis, colon cancer, leukemia, AIDS, thalamus disorders, metabolic disorders including diabetes and obesity, lung diseases such as asthma, emphysema, cystic fibrosis, cancer, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer. The NOV8 nucleic acid encoding B7 family-like protein, and the B7 family-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV9

NOV9 includes four novel Acyl CoA Dehydrogenase-like proteins disclosed below. The disclosed proteins have been named NOV9a, NOV9b, NOV9c and NOV9d.

NOV9a

A disclosed NOV9a nucleic acid of 1446 nucleotides (also referred to cg-140509446) encoding a novel Acyl-CoA Dehydrogenase-like protein is shown in Table 9A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 88-90 and ending with a TAG codon at nucleotides 1444-1446. A putative untranslated region upstream from the initiation codon is underlined in Table 9A, and the start and stop codons are in bold letters.

Table 9A. NOV9a Nucleotide Sequence (SEQ ID NO:35)

CGCGAACAACTGGAAGCTGACCGAATTGTGTCTAACCTGGCGTGGGATTTCGCAGTCAAAGAAGGGTTCCGGGTTTT
CAAAGAGATGCCCTTCAACAATCCGTTAACAAGGTCTACACACGTGGGCCAGGCCCCAGTCCCAGTGGTGGCCACAG
GCAGCAGGAGTTATAGTCCGTTCCAGAAGCTTCCCGAGCTCATACCTCAAGGGGAGGTCTGGTTATCTCTCCAGAGAGC
CTCTCTCCACCTGTGAGAGAGCTGTATCACCAGGTGAAGCACTTCATGGAGCAACGTGTGTACCTGCAGAGCCAGAGCT
GCAGAGTCAACAGGCCTCAGCAGCCAGGTGGAGCCCTCCCGACTGATCGAAGACCTCAAGGAGAAAAGCCAAAGCTGAAG
GACTTTGGAACCTTTTCTACCTTAGAGGCTGATCCCGAGAAAAAATACGGAGCAGGACTGACCAATGTGGAATATGCA
CATCTGTGTGAGCTCATGGGCACGTCCCTGTATGCCCGGAGGTATGTAACCTGCTGCGCCTGACACGGGCAACATGGA
GCTGCTGGTGAGGTATGGCACCAGCGCAGAGGCTCGTGGCTGATTCTCTGCTGGAGGGGAAAGCCCGCTCTGT
TTGCTATGACCCGAGCCAGGTTCCTCTTCCAGATGCCACCAACATTGAGGCTTCATCAGAGAGGAGGACAGCTTCTAT
GTCATAAACGGTCACAAATGGTGGATCAGGCATCCTGGATCTCTGTTGCCAATCTGTGTGTTTATGGGAAAAACAGA
CCCAGATGCACCAAGACACCGGCAGCAGTCTGTGCTCTTGGTTCCTATGGATACCCAGGGGATAAAAATCATCCGGCCTC
TGACGGTGTATGAGTGGAGATGCACAGGTGGCCATGGTGAAGTCCGATTGAGCACGTGCGTGTGCCCAAAGAGAAC
ATGGTCTTGGCCCTGGCCGAGGCTTTGAGATCGCCAGGCGAGACTGGGCCCGGCAGGATCCATCACTGCATGAGGCT
GATCGGGTCTCAGAGAGGGCCCTGGCACTCATGAAGGCCCGCTGAAGTCCCGCTTGGCTTTTGGGAAGCCCGTGGTGG
AGCAGGGCACAGTGTGGCGGACATCGCGCAGTCCGCGGTGGAGATTGAGCAGGCACGGCTGCTGGTGTGAGAGCTGCC
CACCTCATGGACCTGGCAGGAAACAAGGCTGCAGCCTTGGATATAGCCATGATTAATGTTGCGCCCGTCCATGGCCTC
CCGAGTGATTGATCGTGCATTACGGCCTTTGGAGCAGCAGGCTGAGCAGCGACTACCACTGGCTCAGTTCTTACCT
GGGCCCGAGCCCTGCGCTTTGCCGACGGCCCTGACGAGGTGCACGGGCCACGGTGGCCAAGCTAGAGCTGAAGCACC
ATTAG

The disclosed NOV9a nucleic acid sequence, maps to chromosome 12, has 236 of 360 bases (65%) identical to a *Caenorhabditis elegans* cosmid K09H11 mRNA from (gb:GENBANK-ID:CELK09H11|acc:U97002.2) ($E = 2.8e^{-17}$).

A disclosed NOV9a polypeptide (SEQ ID NO:36) encoded by SEQ ID NO:35 is 452 amino acid residues and is presented using the one-letter amino acid code in Table 9B. Signal P, Psort and/or Hydropathy results predict that NOV9a does not contain a signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500.

Table 9B. Encoded NOV9a protein sequence (SEQ ID NO:36).

MPFTNPLTRSYHTWARFQSQWCPGTSRSYSSVPEASPAHSTRGGLVISPESLSPPVRELYHRLKHFMQRVYPAPPELQS
HQASAAARWSPSLIEDLKEKAKAEGWNLFLPLEADPEKKYAGLNTVEYAHLCMLMGTSLYAPEVCNCSAPDTGMELL
VRYGTEAQKARWLIPLLEGKARSCFAMTEPQVASSDATNIEASIREEDSFYVINGHKWITGILDPRCQLCVFMGKTDPH
APRHQQSVLLVPMPTPGIKIRPLTVYGLEDPGGHGEVRFHVVRPKENMVLGPGRGFEIAQGRLGPGRIHHCMLRIG
FSERALALMKARVKSRLAFGKPLVEQGTVLADIAQSRVEIEQARLLVLRAAHLMDLAGNKAALDIAMIKMVAPSMASRV
IDRAIQAFGAAGLSSDYPLAQFFTWARALRFADGPDEVHRATVAKLELKHRI

The NOV9a amino acid sequence has 245 of 396 amino acid residues (61%) identical to, and 294 of 396 amino acid residues (74%) similar to, the *Pseudomonas aeruginosa* 409 amino acid residue probable acyl-CoA dehydrogenase (ptrn:TREMBLNEW-ACC:AAG05938) ($E = 6.4e^{-129}$).

NOV9a is expressed in at least the following tissues: Adipose, Amygdala, Brain, Colon, Foreskin, Hair Follicles, Heart, Kidney, Liver, Ovary, Parathyroid Gland, Pituitary Gland, Placenta, Prostate, Stomach, Thymus, Thyroid, Tonsils and Uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV9b

A disclosed NOV9b nucleic acid of 1363 nucleotides (also referred to CG55900-02) encoding a novel Acyl-CoA Dehydrogenase-like protein is shown in Table 9C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 5-7 and ending with a TAG codon at nucleotides 1361-1363. A putative untranslated region upstream from the initiation codon is underlined in Table 9C, and the start and stop codons are in bold letters.

Table 9C. NOV9b Nucleotide Sequence (SEQ ID NO:37)

AGAGATGCCCTTCACAAATCCGTTAACAAGGTCCTACCACAGTGGGCCAGGCCCCAGTCCCAGTGGTGCCCCACAGGCA
GCAGGAGTTATAGCTCCGTTCCAGAAGCTTCCCCAGCTCATACCTCAAGGGGAGGTCTGGTTATCTCTCCAGAGAGCCTC
TCTCCACCTGTGAGAGAGCTGTATCACCGGCTGAAGCACTTCATGGAGCAACGTGTGTACCTGCAGAGCCAGAGCTGCA
GAGTCACAGGCCCTCAGCAGCCAGGTGGAGCCCTCCCCACTGATCGAAGACCTCAAGGAGAAAGCCAAAGCTGAAGGAC
TTTGGAACCTTTTCTTACCCTTAGAGGCTGATCCCCAGAAAAATACGGAGCAGGACTGACCAATGTGGAATATGCACAT
CTGTGTGAGCTCATGGGCAGTCCCTGTATGCCCGGAGGTATGTAAGTCTCTGCGCCTGACACGGGCAACATGGAGCT
GCTGGTGAGGTATGGCACCAGGAGCGCAGAAGGCTCGTGGCTGATTCTCTGCTGGAGGGAAAGCCGCTCCTGTTTTG
CTATGACCGAGGCCAGGTTGCCTCTTCAGATGCCACCAACATTGAGGCTTCCATCAGAGAGGAGGACAGCTTCTATGTC
ATAAACGGTCACAAATGGTGGATCACAGGCATCCTGGATCCTCGTTGCCAACTCTGTGTGTTTATGGGAAAAACAGACCC
ACATGCACCAAGACACCGGCAGCAGTCTGTGCTCTTGGTTCCCATGGATACCCAGGGATAAAAAATCATCCGGCCTCTGA
CGGTGTATGGACTGGAAGATGCACAGGTGGCCATGGTGAAGTCCGATTGAGCACGTGCGTGTGCCCAAAGAGAACATG
GTCTTGGGCCCTGGCCGAGGCTTTGAGATCGCCAGGGCAGACTGGGCCCGGAGGATCCATCACTGCATGAGGCTGAT
CGGGTCTCAGAGAGGGCCCTGGCACTCATGAAGGCCCGCGTGAAGTCCCGCTTGGCTTTTGGGAAGCCCTGGTGGAGC
AGGGCACAGTGTGCGGACATCGCGCAGTCCGCGTGGAGATTGAGCAGGCACGGCTGCTGGTGTGAGAGCTGCCAC
CTCATGGACCTGGCAGGAAACAAGGCTGCAGCCTTGGATATAGCCATGATTAAAAATGGTCGCCCGCTCCATGGCCTCCCG
AGTGATTGATCGTGGCAATTGAGGCTTTGGAGCAGCAGGCTGAGCAGCGACTACCCACTGGCTCAGTTCTTCACTGGG
CCCGAGCCCTGCGCTTTGCCGACGCGCTGACGAGGTGCACCGACACCGGTGGCCAAGCTAGAGCTGAAGCACCGCATT
TAG

The disclosed NOV9b nucleic acid sequence, maps to chromosome 12, has 778 of 1177 bases (66%) identical to a *Deinococcus radiodurans* mRNA for radiodurans R1 section 1 of 2 of the complete chromosome 2 (gb:GENBANK-ID:AE001862|acc:AE001862.1) ($E = 7.0e^{-84}$).

A disclosed NOV9b polypeptide (SEQ ID NO:38) encoded by SEQ ID NO:37 is 452 amino acid residues and is presented using the one-letter amino acid code in Table 9D. Signal P, Psort and/or Hydropathy results predict that NOV9b does not contain a signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500.

Table 9D. Encoded NOV9b protein sequence (SEQ ID NO:38).

MPFTNPLTRSYHTWARPQSQWCPTGSRYSYSSVPEASPAHTSRGGLVISPESLSPVPVRELYHRLKHFMEQRYVPAEPQLQS
HQASAAWSPSPLEIDLKEKAKAEGWLNLFLPLEADPEKKYAGLTNVEYAHLCELMGTSLYAPEVCNCSAPDGTGMELL
VRYGTEAQKARWLIPLLEGKARSCFAMTEPQVASSDATNIEASIREEDSFYVINGHKWWTGILDPRCQLCVFMGKTDPH
APRRHQQSVLLVPMDTPGIKIIRPLTVYGLEDAPGGHGEVRFHVVRPKENMVLGPGRGFEIAQGRGLGPGRIHHCMRLIG
FSERALALMKARVKSRLAFGKPLVEQGTVLADIAQSRVEIEQARLLVLRRAHLMDLAGNKAAALDIAMIKMVAPSMASRV
IDRAIQAFGAAGLSSDYPLAQFFTWARALRFADGPDVHRRTTAKLELKHRI

The NOV9b amino acid sequence has 247 of 399 amino acid residues (61%) identical to, and 296 of 399 amino acid residues (74%) similar to, the *Deinococcus radiodurans* 415 amino acid residue Acyl-CoA dehydrogenase, putative (ptnr:SPTREMBL-ACC:Q9RYW0) ($E = 1.2e^{-129}$).

NOV9b is expressed in at least the following tissues: Adipose, Adrenal Gland/Suprarenal gland, Amnion, Amygdala, Aorta, Brain, Cervix, Colon, Foreskin, Hair Follicles, Heart, Kidney, Liver, Lung, Ovary, Parathyroid Gland, Pituitary Gland, Prostate, Retina, Right Cerebellum, Stomach, Thymus, Thyroid, Tonsils and Uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention.

NOV9c

A disclosed NOV9c nucleic acid of 1380 nucleotides (also referred to CG55900-03) encoding a novel Acyl-CoA Dehydrogenase-like protein is shown in Table 9E. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 88-90 and ending with a TAG codon at nucleotides 1300-1302. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 9E, and the start and stop codons are in bold letters.

Table 9E. NOV9c Nucleotide Sequence (SEQ ID NO:39)

GCGGAACAACTGGAAAGCTGACCGAATTGTGTCTAACCTGGCGTGGGATTTCGCAGTCAAAGAAGGGTTCCGGGTTTT
 CAAAGAGATGCCCTTCACAAATCCGTTAACAAGGTCTACACACGTGGGCCAGGCCAGTCCCAGTGGTCCCCACAG
 GCAGCAGGAGTTATAGCTCCGTTCCAGAAGCTTCCCAGCTCATACCTCAAGGGGAGGTCTGGTTATCTCTCCAGAGAGC
 CTCTCTCCACCTGTCCAGAGAGCTGTATCACCAGGCTGAAGCACTTCATGGAGCAACGTGTGTACCTGCAGAGCCAGAGCT
 GCAGAGTCACCGAGCCTCAGCAGCCAGGTGGAGCCCTCCCCACTGATCGAAGACCTCAAGGAGAAAGCCAAAGCTGAAG
 GACTTTGGAACCTTTCTACCTTAGAGGCTGATCCCGAGAAAAAATACGGAGCAGGACTGACCAATGTGGAATATGCA
 CATCTGTGTAGCTCATGGGCACGTCCCTGTATGCCCCGAGGTATGTAAGTGTCTGCGCCTGACACGGCAACATGGA
 GCTGCTGGTGAGGTATGGCACCGAAGCGCAGAAAGGTCTGCTGGCTGATTCTCTGCTGGAGGGGAAAGCCCGCTCTGT
 TTGCTATGACCGAGCCCGAGGTGGCTCTTCAGATGCCACCAACATTGAGGCTTCCATCAGAGAGGAGGACAGCTTCTAT
 GTCATAAACCGTCAACAAATGGTGGATCAGGCATCTGGATCTCGTTGCCAATCTGTGTGTTATGGGAAAAACAGA
 CCCACATGCACCAAGACACCGGCAGCAGTCTGTGCTCGTGGTTCCTATGGATACCCAGGGATAAAAAATCATCCGGCTC
 TGACGGTGTATGGACTGGAAGATGCACCGAGGTGGCCATGGTGAAGTCCGATTGAGCACGTGCGTGTGCCAAAGAGAAC
 ATGGTCTGGGCGCTGGCCGAGGCTTTGAGATCGCCAGGCGAGTGGGCCCCGGCAGGATCCATCACTGCATGAGGCT
 GATCGGGTGTGAGAGGGCCATGGCACTCATGAAGGCCCGCGCTGCAGCATTGGATATAGCCATGATTAAAAATGGTCTG
 CCCCCTCCATGGCCTCCCGAGTGATTGATCGTGCATTGAGGCTTTGGAGCAGCAGGCTTGAGCAGCGAATACCCACTG
 GCTCATTTTTCATGGGCCCCGAGCCTTGGCGTCTGACGAGGTGACCGGGCCACGGTGGCCAAAGCT
 AGAGCTGAAGCACCGCATTAGAGCCTTGGGGCTGCAGTGGCTCAATGTCTTGGTGGTCCAGCTGTGCCAGATCTGTCT
 ACTGATGTGCTCGAAAGAT

The disclosed NOV9c nucleic acid sequence, maps to chromosome 12, has 1085 of 1090 bases (99%) identical to a *Homo sapiens* clone MGC:5601 mRNA (gb:GENBANK-ID:BC003698|acc:BC003698.1) ($E = 1.5e^{-253}$).

A disclosed NOV9c polypeptide (SEQ ID NO:40) encoded by SEQ ID NO:39 is 404 amino acid residues and is presented using the one-letter amino acid code in Table 9F. Signal P, Psort and/or Hydropathy results predict that NOV9c does not contain a signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500.

Table 9F. Encoded NOV9c protein sequence (SEQ ID NO:40).

MPFTNPLTRSYHTWARPQSQWCPTGSRYSYSSVPEASPAHSTRGGLVISPESLSPPVRELYHRLKHFMEQRVYPPEPELOS
 HQASARWSPSLIEDLKEKAKAEGLNFLPLEADPEKKYGAGLTNVEYAHLCELMGTSLYAPEVCNCSAPDTGNMELL
 VRYGTEAQKARWLIPLLEGKARSCFAMTEPQVASSDATNIEASIREEDSFYVINGHKWWTGILDPRCOLCVFMGKTDPH

APRRHQSVLVVPMDDPGIKIRPLTVYGLEADPGGHGEVRFHVVRPKENMVLGPGRGFETIAQGRLLGPRIHHCMRLIG
LSERAMALMKARAAALDIAMIKMVAPSMASRVIDRAIQAFGAAGLSSEYPLAHFFTWARALRFADGPDEVHRATVAKLEL
KHRI

The NOV9c amino acid sequence has 263 of 266 amino acid residues (98%) identical to, and 265 of 266 amino acid residues (99%) similar to, the *Homo sapiens* 455 amino acid residue MGC:5601 protein (ptnr:TREMBLNEW-ACC:AAH03698) ($E = 3.1e^{-144}$).

- 5 NOV9c is expressed in at least the following tissues: Adipose, Adrenal Gland/Suprarenal gland, Amnion, Amygdala, Aorta, Brain, Cervix, Colon, Foreskin, Hair Follicles, Heart, Kidney, Liver, Lung, Ovary, Parathyroid Gland, Pituitary Gland, Prostate, Retina, Right Cerebellum, Stomach, Thymus, Thyroid, Tonsils and Uterus. This information was derived by determining the tissue sources of the sequences that were included in the
- 10 invention.

NOV9d

- A disclosed NOV9d nucleic acid of 3490 nucleotides (also referred to CG55900-04) encoding a novel Acyl-CoA Dehydrogenase-like protein is shown in Table 9G. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 113-115 and ending with a TAA codon at nucleotides 3353-3355. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 9G, and the start and stop codons are in bold letters.
- 15

Table 9G. NOV9d Nucleotide Sequence (SEQ ID NO:41)

TCCCATGGGTGTGTAGAAACAGCCTTCCTGAAACACACCCAGCGCAGGCACCAGGGGTCCACCGATGGACACACCTTG
GAGGCAGCACCTACAGAGCGGTGATTTTCGACATGGGCGGAGTTCTCATTCTCTCCAGGGAGAGTCGCTGCAGAATGG
GAGGTACAGAATCGTATCCCTTCCTGGAACATATATAAGGCCCTTGATGGAAGGTGGTGAAATGGGCCCTGGATGAGATT
TATGAGAGCAGAAATAACAGCAGAGGGTTTTTACGAGAATTTGGGAGACTTGTCTGAAATGTTAAAGACCTCCGTGC
CTGTGGACTCATTTTTCTCTCTGTGACCACTGAGCGAGTGGCAAAGCAGTCCCAGTGATGACTGAGGCCATAACTCAA
ATTCCGGGCAAAAGGCTTCAGACTGCAGTCTTGAGCAATAATTTTATCTTCCCAACAGAAAAGCTTTTGGCCCTGGGA
CCGGAACACAGTTTGTATGTGATTGTGGAGTCTGATGGAAGGGATCTGTAAGCCAGACCCTAGGATCTACAAGCTGTGCT
TGGAGCAGCTCGGCCCTGCAGCCCTCTGAGTCCATCTTCTGTATGACCTTGAACAAATCTAAAGAAGCTGCCAGACTT
GGTATTCACACCATTAAGGTTAATGACCCAGAGACTGCAGTAAAGGAATTAGAAGCTCTCTGGGTTTTACATTGAGAGT
AGGTGTTCCAAACACTCGGCCCTGTGAAAAGACGATGGAAATTCGAAAGATTCTTGCAAGTACCTCAAAGACTTAC
TGGGTATCCAGACCACAGGCCCATTCGAACTACTTCAGTTTGATCACGGGCAGTCAAATCCAATCTACTACATCAGGCTG
GCTAATCGTGATCTAGTTCTGAGGAAGAAGCCCCAGGGACACTCCTTCCATCTGCCCATGAGAGAGGGAGTTTCAG
GATTATGAAAGCCCTTGCAATGCTGGAGTACCTGCCCCTAACGTTCTTGATCTCTGTGAAGATTCAAGTGTCTATTGGCA
CCCCCTTCTATGTGATGGAGTACTGCCAGGTCTCATCTACAAAGACCTTCCCTGCCAGGCTTGGAGCCAGCCACAGA
CGAGCCATATACACTGCCATGAACACAGTCTGTGCAAAATTCACAGTGTGGATCTGCAGGCTGTGGGACTTGAAGACTA
TGGGAAGCAAGGGGACTATATTCACGCGCAGGTACGAACCTGGGTAAAGCAGTATCGAGCTTCCGAACTAGCACCATCC
CAGCCATGGAGAGGCTGATCGAATGGCTGCCCTCCATCTTCCCGTCAGCAGAGGACCACAGTGGTGACGGGGACTTC
AGGCTCGACAACCTGGTGTTCATCCAGAAGAGCCAGAGGTGCTTGTCTGCTGACTGGGAACCTTCTACCTTGGGCGA
CCCCCTTGTGTATGTGGCTTACAGTGCCTGGCTCATTAACCTGCCATCCAGTTTCCCGTGTGAGAGGATTAATGACT
GTGACTTGACACAGCTGGGAATCCCTGCTGCAGAGGAGTATTTAGGATGTAAGTCTCTCAAATGGGGCTCCCTCCACT
GAGAAGTGAACCTCTATATGGCTTTTCTCTTTTCCGTGTGGCTGCAATCTACAGGAGTCTACAAGCGATCACTCAC
AGGGAAGCAAGCCCTCCACATATGCGGAACAACTGGAAGCTGACCGAATTTGTGTCTAACCTGGCGTGGGATTTCGCAG
TCAAAGAAGGGTTCCGGGTTTTCAAAGAGATGCCCTTCACAAATCCGTTAAAGAGTCTTACCACAGTGGGCCAGGGCC
CAGTCCCAGTGGTGGCCATAGGCAGCAGGAGTTATAGCTCCGTTCAGAAGCTTCCCGAGTCTATACCTCAAGGGGAGG
TCTGGTTATCTTCCAGAGAGCCTCTCTCCACCTGTGAGAGAGTGTATCACCGGCTGAAGCACTTCATGGAGCAACGTG
TGTACCTTCAGAGCCAGAGCTGCAGAGTCAACAGGCCCTCAGCAGCCAGGTGGAGCCCTCCCACTGATCGAAGACCTC
AAGGAGAAAGCCAAAGCTGAAGGACTTTTGAACCTTTCTACCTTAGAGGCTGATCCCGAGAAAAAATACGGAGCAGG
ACTGACCAATGTGGAATATGCACATCTGTGTGAGCTCATGGGCAGTCCCTGTATGCCCCGAGGTATGTAAGTCTCTG
CGCTGACACGGGCAACATGGAGCTGCTGGTGGGTATGGCACCAGGCGAGAGGCTCGCTGGCTGATTCTCTGCTG
GAGGGGAAAGCCGCTCCTGTTTGCTATGACCGAGCCCAAGGTTGCCTCTTCAGATGCCACCAACATTGAGGCTCCAT

CAGAGAGGAGGACAGCTTCTATGTCATAAACGGTCACAAATGGTGGATCACAGGCATCCTGGATCCTCGTTGCCAACTCT
GTGTGTTTATGGGAAAACAGACCCACATGCACCAAGACACCGGCAGCAGTCTGTGCTCTTGGTTCCCATGGATACCCCA
GGGATAAAAATCATCCGGCTCTGACGGTGTATGGACTGGAAGATGCACAGGTGGCCATGGTGAAGTCCGATTGAGCA
CGTGGCTGTGCCCAAAGAGAACATGGTCTGGGCCCTGGCCGAGGCTTTGAGATCGCCAGGGCAGACTGGGCCCCGGCA
GGATCCATCACTGCATGAGGCTGATCGGGTTCTCAGAGAGGGCCCTGGCACTCATGAAGGCCCGCTGAAGTCCCGCTTG
GCTTTTGGGAAGCCCCCTGGTGGAGCAGGACAGTGTGGCGGACATCGCACAGTCCGCGCTGGAGATTGAGCAGGCACG
GCTGCTGGTGTGAGAGCTGCCACCTCATGGACCTGGCAGGAAACAAGGCTGCAGCCTTGGATATAGCCATGATTA
TGGTTCGCCCCGTCCATGGCTCCCGAGTGATTGATCGTGCATTGAGAGACGCTCTGCAAGAAGCCCTGGAGTCTGTTT
CAGGCCAGGAGGAGGGGATTGCTGAGGGCCAGGGGTTCTGGGACAGAGTCTGGAAAGCTGGTCTTCAGGCTCTCAGT
CCCAGGCTGGGCAGGCACGGTCACTTCACTTTCAGCCTTTTCACTCCCTCTCTCTCTGCTGTGGGAATCTGGACACATTTT
GGGAGGCTCCCAAGGCTGTGGGACGTGCTTGTCTGGCAGCTGCAGGTTCTGTCTGGCCTCCCTGGTGGCAGAGGG
GGCGCCACGGCGGTTGGTGGCTAGAGACCCAGGACCTGGCGCCTGGGAAAATGGAATGCAACCCACATTGTAAGCCA
CTGGCATCTGATTATCTCCATTTGAACACACAGCACAGAACATCATTTAAATGTTATTTTGGAAAGGGGTTTGGGGAC
ACAGAAGAATAAGTAAACACATCTCGGAGGCCAAAAA

The disclosed NOV9d nucleic acid sequence, maps to chromosome 12q24, has 2878 of 2879 bases (99%) identical to a *Homo sapiens* clone MGC:5601 mRNA (gb:GENBANK-ID:BC003698|acc:BC003698.1) (E = 0.0).

A disclosed NOV9d polypeptide (SEQ ID NO:42) encoded by SEQ ID NO:41 is 1080 amino acid residues and is presented using the one-letter amino acid code in Table 9H. Signal P, Psort and/or Hydropathy results predict that NOV9d does not contain a signal peptide and is likely to be localized in the microbody (peroxisome) with a certainty of 0.3930.

Table 9H. Encoded NOV9d protein sequence (SEQ ID NO:42).

MGGVLIIPSPGRVAAEWEVQNRIPSGTILKALMEGGENGPMRFRABEITAEGLREFGRILCSEMLKTSVPVDSFFPSLLTS
ERVAKQFPVMTAEITQIRAKGLQTAFLSNNFYLPNQKSLPLDRKQFDVIVESCMEGICKPDPRIYKLCLEQLGLQPSSES
IFLDDLGTNLKEAARLGIHTIKVNDPETAVKELEALLGFTLRVGVNTRPVKKTMEIPKDSLQKYLKDLLGIQTGTPLEL
LQFDHQSNPTYYIRLANRDLVLRKKPPGTTLLPSAHAIEREFRIMKALANAGVPVNPVLDLCESSVIGTPFYVMEYCPG
LIYKDPSPGLPEPSHRRAIYTAMNTVLCKIHSVDLQAVGLQVLEQYKQGDYIPQVVRTWVKYRASETSTIPAMERLIEWLP
LHLPRQQRRTTVVHGDFRLDNLVFHPPEEVLAVLDWELSTLGDPLADVAYSCLAHYLPSSFPVLRGINDCDLTQLGIPAA
EEYFRMYCLQMGLPPTENWNFYMAFSFFRVAAIILQGVYKRSLTGQASSTYAEQTKLTFEVSNLAWDFAVKEGFRVFKEM
PFTNPLTRSHTWARPQSQWCPIGRSYSSVPEASPAHTSRGGLVISPESLSPVRELYHRLKHFMEQRVYPAEPELQSH
QASAAWSPSPLEIDLKEKAKAEGLNWLFLEADPEKKYAGLTNVEYAHLCMLGTSYAEVCNCSAPDTGNMELLV
RYGTEAQKARWLIPLLEGKARSCFAMTEPQVASSDATNIEASIREDSFYVINGHKWITGILDPRCQLCVFMGKTDPHA
PRHRQOSVLLVPMPTPGIKIIRPLTVYGLEADPGHGEVRFHVVRVPKENMVLGPGRGFETIAQGRLLGPGRIGRHHCMRLIGF
SERALALMKARVKSRLAFGKPLVEQGTVLADIAQSRVEIEQARLLVLRRAHMLDLAGNKAALDIAMIKMVAPSMASRV
DRAIQKTSLQEAWSLFQARRRGFAEQCGSGTESGKLVFRLSVPGWAGTVTSLQPFSPSLSACGNLDTFWEASQCGCTCL
LWQLQGSCLASLVSRGAATAGGGLETQDLGAWENGMOPTL

The NOV9d amino acid sequence has 455 of 577 amino acid residues (78%) identical to, and 503 of 577 amino acid residues (87%) similar to, the *Mus musculus* 629 amino acid residue 2410021P16RIK protein (ptnr:SPTREMBL-ACC:Q9CRH8) (E = 3.5e⁻²⁴⁸).

NOV9d is expressed in at least the following tissues: Mammalian Tissue, Salivary Glands, Liver and Mammary gland/Breast. This information was derived by determining the tissue sources of the sequences that were included in the invention.

Possible small nucleotide polymorphisms (SNPs) found for NOV9b is listed in Table 9I.

Table 9I: SNPs			
Consensus Position	Depth	Base Change	PAF
161	30	C > T	0.167

NOV9a, NOV9b, NOV9c and NOV9d are very closely homologous as is shown in the amino acid alignment in Table 9J.

Table 9J Amino Acid Alignment of NOV9a - NOV9d

	10	20	30	40	50	60	70
NOV9a
NOV9b
NOV9c
NOV9d	MGGVLI	I P S P G R V A A E W E V Q N R I P S G T I L K A L M E G G E N G P W M R F M R A E I T A E G F L R E F G R L C S E M L K T S V P					
	80	90	100	110	120	130	140
NOV9a
NOV9b
NOV9c
NOV9d	V D S F F S L T S E R V A K Q F P V M T E A I T Q I R A K G L Q T A V L S N N F Y L P N Q K S F L P L D R K Q F D V I V E S C M E G I C K						
	150	160	170	180	190	200	210
NOV9a
NOV9b
NOV9c
NOV9d	P D P R I Y K L C L E Q L G L Q P S E S I F L D D L G T N L K E A A R L G I H T I K V N D P E T A V K E L E A L L G F T L R V G V P N T R P						
	220	230	240	250	260	270	280
NOV9a
NOV9b
NOV9c
NOV9d	V K K T M E I P K D S L Q K Y L K D L L G I Q T T G P L E L L Q F D H G Q S N P T Y Y I R L A N R D L V L R K K P P G T L L P S A H A I E R						
	290	300	310	320	330	340	350
NOV9a
NOV9b
NOV9c
NOV9d	E F R I M K A I A N A G V P V P N V L D L C E D S S V I G T P F Y V M E Y C P G L I Y K D P S L P G L E P S H R R A I Y T A M N T V L C K I						
	360	370	380	390	400	410	420
NOV9a
NOV9b
NOV9c
NOV9d	H S V D L Q A V G L E D Y G K Q G D Y I P R Q V R T W V K Q Y R A S E T S T I P A M E R L I E W L P L H L P R Q Q R T T V V H G D F R L D N						
	430	440	450	460	470	480	490
NOV9a
NOV9b
NOV9c
NOV9d	L V F H P E E P V L A V L D W E L S T L G D P L A D V A Y S C L A H Y L P S S F P V L R G I N D C D L T Q L G I P A A E E Y F R M Y C L Q						
	500	510	520	530	540	550	560
NOV9a
NOV9b
NOV9c
NOV9d	M G L P P T E N W N F Y M A F S F F R V A A I L Q G V Y K R S L T G Q A S S T Y A E Q T G K L T E F V S N L A W D F A V K E G F R V F K E M						
	570	580	590	600	610	620	630
NOV9a	P F T N P L T R S Y H T W A R P Q S Q W C P T G S R S Y S S V P E A S P A H T S R G G L V I S P E S L S P P V R E L Y H R L K H F M E Q R V						
NOV9b	P F T N P L T R S Y H T W A R P Q S Q W C P T G S R S Y S S V P E A S P A H T S R G G L V I S P E S L S P P V R E L Y H R L K H F M E Q R V						
NOV9c	P F T N P L T R S Y H T W A R P Q S Q W C P T G S R S Y S S V P E A S P A H T S R G G L V I S P E S L S P P V R E L Y H R L K H F M E Q R V						
NOV9d	P F T N P L T R S Y H T W A R P Q S Q W C P T G S R S Y S S V P E A S P A H T S R G G L V I S P E S L S P P V R E L Y H R L K H F M E Q R V						
	640	650	660	670	680	690	700
NOV9a	Y P A E P E L Q S H Q A S A A R W S P S P L I E D L K E K A K A E G L W N L F L P L E A D P E K K Y G A G L T N V E Y A H L C E L M G T S L						
NOV9b	Y P A E P E L Q S H Q A S A A R W S P S P L I E D L K E K A K A E G L W N L F L P L E A D P E K K Y G A G L T N V E Y A H L C E L M G T S L						
NOV9c	Y P A E P E L Q S H Q A S A A R W S P S P L I E D L K E K A K A E G L W N L F L P L E A D P E K K Y G A G L T N V E Y A H L C E L M G T S L						
NOV9d	Y P A E P E L Q S H Q A S A A R W S P S P L I E D L K E K A K A E G L W N L F L P L E A D P E K K Y G A G L T N V E Y A H L C E L M G T S L						

710 720 730 740 750 760 770

NOV9a YAEVNCNCSAPDTGNMELLVRYGTEAQKARWLIPLLEGKARSCFAMTEPQVASSDATNIEASIREEDSFY

NOV9b YAEVNCNCSAPDTGNMELLVRYGTEAQKARWLIPLLEGKARSCFAMTEPQVASSDATNIEASIREEDSFY

NOV9c YAEVNCNCSAPDTGNMELLVRYGTEAQKARWLIPLLEGKARSCFAMTEPQVASSDATNIEASIREEDSFY

NOV9d YAEVNCNCSAPDTGNMELLVRYGTEAQKARWLIPLLEGKARSCFAMTEPQVASSDATNIEASIREEDSFY

780 790 800 810 820 830 840

NOV9a VINGHKWWTGILDPRCQLCVFMGKTDPHAPRHRQOSVLLVPMMDTPGIKIIIRPLTVYGLEDPAGGHGEVR

NOV9b VINGHKWWTGILDPRCQLCVFMGKTDPHAPRHRQOSVLLVPMMDTPGIKIIIRPLTVYGLEDPAGGHGEVR

NOV9c VINGHKWWTGILDPRCQLCVFMGKTDPHAPRHRQOSVLLVPMMDTPGIKIIIRPLTVYGLEDPAGGHGEVR

NOV9d VINGHKWWTGILDPRCQLCVFMGKTDPHAPRHRQOSVLLVPMMDTPGIKIIIRPLTVYGLEDPAGGHGEVR

850 860 870 880 890 900 910

NOV9a FEHVRVPKENMVLGPGRGFEIAQGRLGPGRIHHCMRLIGFSERALALMKARVKSRLAFGKPLVEQGTVLA

NOV9b FEHVRVPKENMVLGPGRGFEIAQGRLGPGRIHHCMRLIGFSERALALMKARVKSRLAFGKPLVEQGTVLA

NOV9c FEHVRVPKENMVLGPGRGFEIAQGRLGPGRIHHCMRLIGFSERALALMKARVKSRLAFGKPLVEQGTVLA

NOV9d FEHVRVPKENMVLGPGRGFEIAQGRLGPGRIHHCMRLIGFSERALALMKARVKSRLAFGKPLVEQGTVLA

920 930 940 950 960 970 980

NOV9a DIAQSRVEIEQARLLVLRRAAHLMDLACNKAALDIAMIKMVAPSMASRVIDRAIQ

NOV9b DIAQSRVEIEQARLLVLRRAAHLMDLACNKAALDIAMIKMVAPSMASRVIDRAIQ

NOV9c DIAQSRVEIEQARLLVLRRAAHLMDLACNKAALDIAMIKMVAPSMASRVIDRAIQ

NOV9d DIAQSRVEIEQARLLVLRRAAHLMDLACNKAALDIAMIKMVAPSMASRVIDRAIQKTSLQEAWSLFOARR

990 1000 1010 1020 1030 1040 1050

NOV9a AFCAAG LSSDYPLAQFFTWARALR FADGPD EVHRTVA

NOV9b AFCAAG LSSDYPLAQFFTWARALR FADGPD EVHRTVA

NOV9c AFCAAG LSSDYPLAQFFTWARALR FADGPD EVHRTVA

NOV9d RGFEECCGSGTSGKLVFRLSVPGWAGTVTSLOPFSPLSACGNLDTFWEASQCGTCLLWQLOGSCLA

1060 1070 1080

NOV9a KLELKHRI

NOV9b KLELKHRI

NOV9c KLELKHRI

NOV9d SLVSRGAATAGGLETQDLGAWENGMQPTL

Homologies to any of the above NOV9 proteins will be shared by the other NOV9 proteins insofar as they are homologous to each other as shown above. Any reference to NOV9 is assumed to refer to both of the NOV9 proteins in general, unless otherwise noted.

- 5 NOV9a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 9K.

Table 9K. BLAST results for NOV9a						
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect	
gi 12846107 dbj BAB 27033.1 (AK010568)	putative [Mus musculus]	629	369/451 (81%)	407/451 (89%)	0.0	
gi 13376864 ref NP_ 079523.1 (NM_025247)	hypothetical protein MGC5601 [Homo sapiens]	455	348/368 (94%)	350/368 (94%)	0.0	
gi 15807862 ref NP_ 285519.1 (NC_001264)	acyl-CoA dehydrogenase, putative [Deinococcus radiodurans]	415	247/399 (61%)	296/399 (73%)	e-139	
gi 15597746 ref NP_ 251240.1 (NC_002516)	probable acyl-CoA dehydrogenase [Pseudomonas aeruginosa]	409	245/396 (61%)	294/396 (73%)	e-136	

gi 15231306 ref NP_187337.1 (NC_003074)	acetyl-coA dehydrogenase, putative [Arabidopsis thaliana]	423	230/394 (58%)	289/394 (72%)	e-131
---	---	-----	------------------	------------------	-------

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 9L.

Table 9L Information for the ClustalW proteins

- 1) NOV9a (SEQ ID NO:36)
- 2) gi|12846107|dbj|BAB27033.1| (AK010568) putative [Mus musculus] (SEQ ID NO:96)
- 3) gi|13376864|ref|NP_079523.1| (NM_025247) hypothetical protein MGC5601 [Homo sapiens] (SEQ ID NO:97)
- 4) gi|15807862|ref|NP_285519.1| (NC_001264) acyl-CoA dehydrogenase, putative [Deinococcus radiodurans] (SEQ ID NO:98)
- 5) gi|15597746|ref|NP_251240.1| (NC_002516) probable acyl-CoA dehydrogenase [Pseudomonas aeruginosa] (SEQ ID NO:99)
- 6) gi|15231306|ref|NP_187337.1| (NC_003074) acetyl-coA dehydrogenase, putative [Arabidopsis thaliana] (SEQ ID NO:100)

	10	20	30	40	50	60	70
NOV9a						
gi 12846107	QAARGTRKGYKCYLQSPFILWFLRLDNLIFHPEKAEVLAVLDWELSTLGDPFADVAYSCLAYLPSSEFP						
gi 13376864						
gi 15807862						
gi 15597746						
gi 15231306						
	80	90	100	110	120	130	140
NOV9a						
gi 12846107	ILRGFRDQDVTKLGIPITVEEYFRMYCLNMGIPPIDNWNFYMAFSFPRVAAILQGVYKRSITGQASSATAQ						
gi 13376864						
gi 15807862						
gi 15597746						
gi 15231306						
	150	160	170	180	190	200	210
NOV9a						
gi 12846107	QSGKLTESMAELAWDFATKEGFRVFKEMPATKILRSYHAWAGPRSPRTPKGVRGHSTVAAASPSHEAKG						
gi 13376864						
gi 15807862						
gi 15597746						
gi 15231306						
	220	230	240	250	260	270	280
NOV9a						
gi 12846107	GLVISPELSPPVRELYHRLGHFMEQRVYPAEPLOSHOASAAARWSESPLIEDLKEKAKAEGLNWLFPL						
gi 13376864	GLVISPEGLSPAARKLYEGLVQETEQKVYLEPELORHCASADRWSESPLIEDLKEKAKAEGLNWLFPL						
gi 15807862						
gi 15597746						
gi 15231306						
	290	300	310	320	330	340	350
NOV9a						
gi 12846107	EADPERKYGAGLTINVEYAHLCCEIMGTSIYAPEVCNCSAPDTGNMELLVRYGTBAOKARWLIPLLECKARS						
gi 13376864	ETDPERKYGAGLTINVEYAHLCCEIMGTSIYAPEVCNCSAPDTGNMELLVRYGTBEOARWLIPLLECKIRS						
gi 15807862	EADPERKYGAGLTINVEYAHLCCEIMGTSIYAPEVCNCSAPDTGNMELLVRYGTBAOKARWLIPLLECKARS						
gi 15597746	ASDPECKEGAGLTINVEYAGLCCEIMGRVWMAPEVENCNAPDTGNMELLVRYGTBEOQEWLIPLLECKIRS						
gi 15231306	---ESEYAGLANHEYAPLADIMCRSLIGPEENCAAPDTGNMELLVRYGSEEOKRTWLEPLLSCBIRS						
	360	370	380	390	400	410	420
NOV9a						
gi 12846107	CFAMTEPOVASSDATNIEASIRREDSHYVINGKWWITGIDPRCOICVFMGKTDPHAPRHRQOSVILVP						

gi|12846107| CFAMTEPOVASSDASNI EASIKEDGCVVINGCHKWWTSCILDPKCLCVFMGKTDFCAPRHQOOSMLVLP
gi|13376864| CFAMTEPOVASSDATNIEASIREEDSRYVINGCHKWWTSCILDPKCLCVFMGKTDPHAPRHRQOOSVLLVP
gi|15807862| AFAMTEPOVASSDATNIEAIVRGDEYVINGRKWSSCAGLSRCKVSTFMGKTDFCAPRHLOOSMLVLP
gi|15597746| AFAMTEPOVASSDATNMEARAERGDDVINGRKWWTSCACDPKCLLTFMCLTFEDAPRHQOOSMLVLP
gi|15231306| GFAMTEPOVASSDATNIECSIRROGDSYVINGCHKWWTSCAMPDPCRVLLFMGKTDFNAPRHQOOSMLVLP

430 440 450 460 470 480 490

NOV9a MDTPGIKITRPLIVVGLDAPCGHGEVRFVFEHVRVPKENMVLGPGRGFEIAQGRLGPGRIHHCMRLIGFSE
gi|12846107| MDSPGITVIRPLSVFGLDEPGCHGEVRFKIDVRVPKENILLGPGRGFEIAQGRLGPGRIHHCMRLIGYSE
gi|13376864| MDTPGIKITRPLIVVGLDAPCGHGEVRFVFEHVRVPKENMVLGPGRGFEIAQGRLGPGRIHHCMRLIGFSE
gi|15807862| FIDAPGVKIERALQVFGFDDAPHGHMMSFDNVRVPVTNMLLGEGRGFEIAQGRLGPGRIHHCMRLVQOAE
gi|15597746| VDTPCVKITRPLIVVFGYDDAPHGHAELVFENVRVPVENVLLGPGRGFEIAQGRLGPGRIHHCMRSIGMAE
gi|15231306| MRTPGISVVRPLIVFGFDDAPHGHAEISFDNVRVPFAKNILLGPGRGFEIAQGRLGPGRIHHCMRLIGAAE

500 510 520 530 540 550 560

NOV9a RALALMKARVKSRLAFGKPLVEQGTVLADIAQSRVEIEQARLLVLRRAAHLMDLAGNKAALDIAMIKMVA
gi|12846107| RALALMKTRVMSRTAFGKPLVEQGTVLADIAQSRVEIEQARLLVLRRAAHLMDVAGNKTAAALDIAMIKMVV
gi|13376864| RALALMKARVKSRLAFGKPLVEQGTVLADIAQSRVEIEQARLLVLRRAAHLMDLAGNKAALDIAMIKMVA
gi|15807862| RALELMIERSACRIAFGKPLALHOHTRELIAQSRMEIDQARLLVLRRAAHTMDTVGNKDAKGEIAAIKVVA
gi|15597746| RALELMIERSACRIAFGKPLARLGNIDHIAQSRMEINMARLLVLRRAAHTMDTVGNKIAQSEIAAIKVVA
gi|15231306| RCMELMAQRALSRKTRCKFLAQHCSFVSDIAKIRVLEEGTRLLVLRRAADHLEKFGNKKARGILAMAKVAA

570 580 590 600 610 620 630

NOV9a PSMASRVIDRAIQ-----AFCAAG-----LSSDYPLAQFETW
gi|12846107| PSMAYHVIDRAIQ-----AFCAAG-----LSSDYPLAQFEGW
gi|13376864| PSMASRVIDRAIQKTLQEAWSLFQARRRGFAEGCGSGTESGKLVFRLVPGWAGTVISLQFSPSLSA
gi|15807862| PNMALRVIDRAIQ-----VYGGAG-----VCODTPLAMMYAQ
gi|15597746| PNMALKVIDRAIQ-----MHCGAG-----VSNDFPLAYWYAM
gi|15231306| PNMALKVIDRAIQ-----VHCAAG-----VSSDTVLAHLWAT

640 650 660 670 680

NOV9a ARALR-FADGPD-----EVHRAIVAKLELKHRI
gi|12846107| ARALR-FADGPD-----EVHOLTIVAKMELKNQS-----R-----MQEPAVPRV
gi|13376864| CGNLDTFWEASCGCGTCLLWQLOGSCLASLVSRGAAATAGGGLETQDLGAWENGMOPTL
gi|15807862| ARTLR-LADGPD-----IVHTEIVAKLELKHRI-----DLRALSKR
gi|15597746| QRTLR-LADGPD-----EVHRAAIGKFELEKYVP-----REMLRSSR
gi|15231306| ARTLR-LADGPD-----EVHGLTIGKLELKHRI-----ASKL

Tables 9M and 9N list the domain description from DOMAIN analysis results against NOV9a. This indicates that the NOV9a sequence has properties similar to those of other proteins known to contain these domains.

Table 9M. Domain Analysis of NOV9a

gnl|Pfam|pfam00441, Acyl-CoA_dh, Acyl-CoA dehydrogenase, C-terminal domain. C-terminal domain of Acyl-CoA dehydrogenase is an all-alpha, four helical up-and-down bundle. (SEQ ID NO:101)
Length = 150 residues, 98.0% aligned
Score = 79.0 bits (193), Expect = 6e-16

NOV9a: 297 GRGFEIAQGRLGPGRIHHCMRLIGFSE RALALMKARVKSRLAFGKPLVEQGTVLADIAQS 356
|+|+|+| | |+ + +|++|| | | | | | + +|
00441: 1 GKGFKYAMKELDMERLVIAAQAALGTAQGAALDEAIPYAKQKQFGKPLAHFQLIQFKLADM 60
NOV9a: 357 RVEIEQARLLVLRRAAHLMDLAGNKAALDIAMIKMVAPSMASRVIDRAIQAFGAAGLSSD 416
++| | | | + | | | | + + + | | + | | + | | +| | +|
00441: 61 ATKLEAARLLLYRAAWLADRG--RPTSKEAAMAKLFASEAAQVADDAVQILGGVGYTND 118
NOV9a: 417 YPLAQFTWARALRFADGPDVEHRAIVAK 445
| | + +|+ | + + +| | + | +|+
00441: 119 YPVERFYRDAKITQIYEGTSEIQRLVIAR 147

Table 9N. Domain Analysis of NOV9a

gnl|Pfam|pfam02770, Acyl-CoA_dh_M, Acyl-CoA dehydrogenase, middle domain. Central domain of Acyl-CoA dehydrogenase has a beta-barrel fold. (SEQ ID NO:102)
Length = 102 residues, 99.0% aligned
Score = 72.4 bits (176), Expect = 5e-14

```
NOV9a: 186 AMTEPQVASSDATNIEASIREEDSFYVINGHKWITGILDPRCQLCVFMGKTDPHAPRHR 245
          ||| | | | +|+ + + |++| | | | + + + ||| | | +
02770: 1 ALTEPG-AGSDVGSIKTTAERKGGDYILNGSKMWITNG--GQADWYIVLAVTDP-APGKK 56

NOV9a: 246 QQSVLLVPM DTPGIKII RPLTVYGLEDAPGGHGEVRF EHV RVPKENM 292
          + || | | | | | + | | + | + | | | | | + | +
02770: 57 GITAPLVEKDTPGFHIGKKEDKGLRSSD--TCELIFEDVRVPESNI 101
```

5 Acyl-CoA is an important energy-storing molecule which can be stored as fat or burned in muscle. Enzymes that modify this molecule may be important obesity and diabetes targets.

Two distinct clinical phenotypes of hereditary short-chain acyl-CoA dehydrogenase (SCAD, or ACADS; EC 1.3.99.2) deficiency have been identified. One type has been
10 observed in infants with acute acidosis and muscle weakness; the other has been observed in middle-aged patients with chronic myopathy. SCAD deficiency is generalized in the former type and localized to skeletal muscles in the latter. Cases with neonatal onset have a variable phenotype that includes metabolic acidosis, failure to thrive, developmental delay, and seizures, as well as myopathy. There are no episodes of nonketotic hypoglycemia, which are
15 characteristic of medium-chain (MCAD; 201450) and long-chain (LCAD; 201460) acyl dehydrogenase deficiencies. The definitive diagnostic test for SCAD deficiency is an ETF-linked enzyme assay with butyryl-CoA as a substrate, performed after immunoactivation of MCAD, which has similar activity (Bhala et al. Clinical and biochemical characterization of short-chain acyl-coenzyme A dehydrogenase deficiency. J Pediatr. 126(6):910-5, 1995; Tein et al., Short-chain acyl-CoA dehydrogenase deficiency: a cause of ophthalmoplegia and
20 multicore myopathy. Neurology. 52(2):366-72, 1999).

Turnbull et al. (N Engl J Med. 311(19):1232-6, 1984) reported the case of a 53-year-old woman who presented with a lipid-storage myopathy and low concentrations of carnitine in skeletal muscle. Impaired fatty acid oxidation in muscle was found to be caused by
25 deficiency of short-chain acyl-CoA (butyryl-CoA) dehydrogenase activity in mitochondria. The authors suggested that the muscle carnitine deficiency was secondary to this enzyme deficiency and urged that it be considered in other cases of lipid-storage myopathy with carnitine deficiency (212160). Onset of myopathy was at age 46 years. Amendt et al. (J Clin Invest. 79(5):1303-9, 1987) described 2 unrelated patients, both of whom presented with

neonatal metabolic acidosis and ethylmalonate excretion. Deficiency of short-chain acyl-CoA dehydrogenase was demonstrated in fibroblasts by both an electron-transfer flavoprotein (ETF)-linked dye-reduction assay and a tritium release ADH assay. The patient described by Turnbull et al. (1984) had normal SCADH activity in fibroblasts, which raises the possibility that a distinct SCADH isoenzyme exists in mammalian muscle. However, Amendt et al. (1992) found that in mice SCAD is the same in both muscle and fibroblasts. For that reason, Bhala et al. (1995) proposed that the case of Turnbull et al. (1984) was not a primary case of SCAD deficiency but rather a case of riboflavin-responsive multiple acyl-CoA dehydrogenase deficiency, as reported by DiDonato et al. (Ann Neurol. 25(5):479-84, 1989).

The protein similarity information, expression pattern, and map location for the NOV9 suggest that NOV9 may have important structural and/or physiological functions characteristic of the Acyl-CoA Dehydrogenase protein family. Therefore, the NOV9 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the NOV9 compositions of the present invention will have efficacy for treatment of patients suffering from obesity, diabetes, cachexia, cancer, inflammation, CNS disorders and SCAD disorders. The NOV9 nucleic acid encoding Acyl-CoA Dehydrogenase-like protein, and the Acyl-CoA Dehydrogenase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring

polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver,

spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

5 A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11,
10 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York,
15 NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore,
20 oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a
25 genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length
30 would further comprise at least 6 contiguous nucleotides SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID

NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of an NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 or 41 is one that is sufficiently complementary to the nucleotide sequence shown NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 or 41 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or

proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See e.g. Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA.

Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, e.g., frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein.

Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

An NOVX polypeptide is encoded by the open reading frame ("ORF") of an NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, e.g., a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning

NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41; or an anti-sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41; or of a naturally occurring mutant of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express an NOVX protein, such as by measuring a level of an NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of an NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, that encodes a polypeptide having an NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein

having an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding an NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

5 As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the
10 thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in
15 which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

20 Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are
25 hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39
30 and 41, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17,

19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.,* Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.,* as employed for cross-species hybridizations). *See, e.g.,* Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without altering the functional ability of said NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX

proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42; more preferably at least about 70% homologous SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42; still more preferably at least about 80% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42; even more preferably at least about 90% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42; and most preferably at least about 95% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42.

An isolated nucleic acid molecule encoding an NOVX protein homologous to the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine,

methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein and an NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules

are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of an NOVX protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42, or
5 antisense nucleic acids complementary to an NOVX nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an NOVX protein. The term
10 "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and
15 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is
20 antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures
25 known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

30 Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine,

2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330).

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for an NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of an NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using

standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S₁ nucleases (*See*, Hyrup, *et al.*, 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*See*, Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see*, Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In

addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g., Krol, et al., 1988. BioTechniques 6:958-976*) or intercalating agents (*see, e.g., Zon, 1988. Pharm. Res. 5: 539-549*). To this end, the oligonucleotide may be conjugated to another molecule, *e.g., a peptide, a hybridization triggered cross-linking agent, a transport agent, a*
5 *hybridization-triggered cleavage agent, and the like.*

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42. The invention also
10 includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42 while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, an NOVX variant that preserves NOVX-like function includes any variant
15 in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as
20 defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an
25 appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof
30 is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated

from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of an NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of an NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42, and retains the functional

activity of the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42, and retains the functional activity of the NOVX proteins of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*,

the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, an NOVX "chimeric protein" or "fusion protein" comprises an NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an NOVX protein SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within an NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of an NOVX protein. In one embodiment, an NOVX fusion protein comprises at least one biologically-active portion of an NOVX protein. In another embodiment, an NOVX fusion protein comprises at least two biologically-active portions of an NOVX protein. In yet another embodiment, an NOVX fusion protein comprises at least three biologically-active portions of an NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is an NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an NOVX ligand and an NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of an NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with an NOVX ligand.

An NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the

biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of an NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded

DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. See, e.g., Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, et al., 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , F_{ab}' and $F_{(ab)2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for

polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX-related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX-related protein sequence will indicate which regions of a NOVX-related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow and Lane, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, *MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE*, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur et al., *MONOCLONAL ANTIBODY PRODUCTION TECHNIQUES AND APPLICATIONS*, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by

the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by
5 limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures
10 such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by
15 using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to
20 obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, *Nature* 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin
25 polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further
30 comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-

binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., *Nature*, 321:522-525 (1986); Riechmann et al., *Nature*, 332:323-327 (1988); Verhoeyen et al., *Science*, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences.

In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)).

Human Antibodies

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon

challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992));
5 Lonberg et al. (*Nature* 368 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild et al, (*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's
10 endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human
15 DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins.
20 The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be
25 further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at
30 least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives; fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F_{(ab')₂} fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_{(ab')₂} fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct

bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen
5 combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the
10 immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh *et al.*, *Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair
15 of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing
20 large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from
25 antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan *et al.*, *Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab'
30 fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another

bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention.

- 5 Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents.
- 10 For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

- 15 It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and
- 20 antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis
- 25 and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

- The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive
- 30 isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-NOVX antibodies may be used in methods known within the art relating to the localization and/or quantitation of an NOVX protein (e.g., for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for

use in imaging the protein, and the like). In a given embodiment, antibodies for NOVX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

5 An anti-NOVX antibody (*e.g.*, monoclonal antibody) can be used to isolate an NOVX polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-NOVX antibody can facilitate the purification of natural NOVX polypeptide from cells and of recombinantly-produced NOVX polypeptide expressed in host cells. Moreover, an anti-NOVX antibody can be used to detect NOVX protein (*e.g.*, in a cellular lysate or cell
10 supernatant) in order to evaluate the abundance and pattern of expression of the NOVX protein. Anti-NOVX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various
15 enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate,
20 rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

NOVX Recombinant Expression Vectors and Host Cells

25 Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA
30 segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*,

non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel,

GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

5 Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the
10 recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors
15 include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

20 Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

25 One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, e.g., Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (see, e.g., Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid
30 sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30:

933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*,
5 SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are
10 often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring
15 Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific
20 promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament
25 promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also
30 encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to

NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.,* DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.,* resistance to antibiotics) is generally introduced into the

host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional

protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene
5 carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. See, e.g., Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is then introduced into
10 an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. See, e.g., Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND
15 EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for
20 constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a
25 system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, See, e.g., Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. See, O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing
30 transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

10 **Pharmaceutical Compositions**

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or

methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, an NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral

antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

5 It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent
10 on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example,
15 intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced
20 intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

25 The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in an NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or
30 expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome

X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease (possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention
5 can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

10 The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

15 In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of an NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid
20 phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

25 A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened
30 with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678;

Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

- Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

- In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to an NOVX protein determined. The cell, for example, can of mammalian origin or a yeast cell.
- Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ¹²⁵I, ³⁵S, ¹⁴C, or ³H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active

portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule. As used herein, a "target molecule" is a molecule with which an NOVX protein
5 binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An NOVX target molecule can be a non-NOVX molecule or an NOVX protein or polypeptide of the invention. In one embodiment, an NOVX target
10 molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

15 Determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by
20 detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+} , diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising an NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation,
25 or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting an NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined
30 either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX

protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to an NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate an NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of an NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of

NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based

upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective
5 genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

10 Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or fragments
15 or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX, sequences can be used to rapidly select primers that do not span more than one exon in the
20 genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals
25 (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be
30 established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. See, e.g., D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human

chromosomes can also be produced by using human chromosomes with translocations and deletions.

5 PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

10 Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique
15 chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see*, Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

20 Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

25 Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage
30 analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then

the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in

SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

5 The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*,
10 blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders,
15 and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in an NOVX gene can be assayed in a
20 biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

 Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or
25 prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

30 Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

 These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

5 In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the
10 presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a
15 standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX
20 expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a
25 disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained
30 from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein,

peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively
5 treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

10 The methods of the invention can also be used to detect genetic lesions in an NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene
15 encoding an NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from an NOVX gene; (ii) an addition of one or more nucleotides to an NOVX gene; (iii) a substitution of one or more nucleotides of an NOVX gene, (iv) a chromosomal rearrangement of an NOVX gene; (v) an alteration in the level of a messenger
20 RNA transcript of an NOVX gene, (vi) aberrant modification of an NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of an NOVX gene, (viii) a non-wild-type level of an NOVX protein, (ix) allelic loss of an NOVX gene, and (x) inappropriate post-translational modification of an NOVX protein. As described herein, there are a large number of assay
25 techniques known in the art which can be used for detecting lesions in an NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a
30 polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see, Abravaya, et al.*, 1995. *Nucl. Acids Res.* 23: 675-682).

This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to an NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see*, Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q β Replicase (*see*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.*, Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al., supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation

array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (see, e.g., Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S₁ nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. See, e.g., Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to

an exemplary embodiment, a probe based on an NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

5 In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.*, Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79.

10 Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in
15 which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.*, Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in
20 polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.*, Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a
25 denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.*, Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer
30 extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.*, Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the

oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g., Gibbs, et al., 1989. Nucl. Acids Res. 17: 2437-2448*) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g., Prossner, 1993. Tibtech. 11: 238*). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g., Gasparini, et al., 1992. Mol. Cell Probes 6: 1*. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g., Barany, 1991. Proc. Natl. Acad. Sci. USA 88: 189*. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

25 Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.,* NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such

treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM

show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

5 Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness
10 phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

15 Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be
20 monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such
25 clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates
30 NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene

expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of an NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS,

bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, and other diseases, disorders and conditions of the like.

These methods of treatment will be discussed more fully, below.

Disease and Disorders

5 Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs,
10 derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination
15 (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

20 Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

25 Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by
30 sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, an NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of an NOVX protein, a peptide, an NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering an NOVX protein or

nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect.

- 5 One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

- 10 In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

- 15 In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

- 20 The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic
25 disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

- As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for
30 treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

Examples

EXAMPLE 1: Identification of NOVX Nucleic Acids

TblastN using CuraGen Corporation's sequence file for polypeptides or homologs was run against the Genomic Daily Files made available by GenBank or from files downloaded from the individual sequencing centers. Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The novel NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. PCR primer sequences were used for obtaining different clones. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain -

hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, 5 cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as 10 components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

Physical clone: Exons were predicted by homology and the intron/exon boundaries 15 were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent 20 inconsistencies thereby obtaining the sequences encoding the full-length protein.

Example 2: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a 25 "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which 30 one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the

region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

5 SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that resulted were selected for further analysis because this identity indicates that these clones contain the
10 genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

15 Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one
20 or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

25 The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn et al., Determination of
30 Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. Genome Research. 10 (8) 1249-1265, 2000).

Example 3. Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System.

5 Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on
10 metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoinflammatory diseases), Panel CNSD.01 (containing samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of
15 agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

20 First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β -actin and GAPDH). Normalized RNA (5 μ l) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

25 In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10 μ g of total RNA were performed in a volume of 20 μ l and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50 μ g of total RNA in a final volume of 100 μ l. sscDNA samples
30 are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a

similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60°C, primer optimal T_m = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe T_m must be 10°C greater than primer T_m , amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

10 PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using
15 TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in
20 RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

 When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers
25 were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

30 The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the

following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

General_screening_panel_v1.4

The plates for Panel 1.4 include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panel 1.4 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer.

Cell lines used in Panel 1.4 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panel 1.4 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or

fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

Panels 2D and 2.2

The plates for Panels 2D and 2.2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum.

These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

Panels 4D, 4R, and 4.1D

5 Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients
10 was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells,
15 microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF
20 alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation,
25 using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2µg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-
30 10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5µg/ml. Samples were

taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2×10^6 cells/ml in DMEM 5% FCS (Hyclone), $100 \mu\text{M}$ non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol ($5.5 \times 10^{-5} \text{M}$) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions.

Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), $100 \mu\text{M}$ non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5 \times 10^{-5} \text{M}$ (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), $100 \mu\text{M}$ non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5 \times 10^{-5} \text{M}$ (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml . Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml . Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at $10 \mu\text{g/ml}$ for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), $100 \mu\text{M}$ non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5 \times 10^{-5} \text{M}$ (Gibco), and 10mM Hepes (Gibco) and plated at 10^6 cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with $0.5 \mu\text{g/ml}$ anti-CD28 (Pharmingen) and $3 \mu\text{g/ml}$ anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), $100 \mu\text{M}$ non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco),

mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5 μ g/ml or anti-CD40 (Pharmingen) at approximately 10 μ g/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24, 48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μ g/ml anti-CD28 (Pharmingen) and 2 μ g/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5 - 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1 μ g/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1 μ g/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 μ g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at 5×10^5 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5×10^5 cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1 μ g/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 μ l of RNase-free water and 35 μ l buffer (Promega) 5 μ l DTT, 7 μ l RNasin and 8 μ l DNase were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at -80°C.

AI_comprehensive_panel_v1.0

The plates for AI_comprehensive_panel_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohn's disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebid and two were on phenobarbital.

Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1 anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI_comprehensive panel_v1.0 panel, the following abbreviations are used:

AI = Autoimmunity

Syn = Synovial

Normal = No apparent disease

Rep22 /Rep20 = individual patients

RA = Rheumatoid arthritis

Backus = From Backus Hospital

OA = Osteoarthritis

(SS) (BA) (MF) = Individual patients

Adj = Adjacent tissue

Match control = adjacent tissues

-M = Male

-F = Female

COPD = Chronic obstructive pulmonary disease

5 Panels 5D and 5I

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample.

- 15 Patient 2: Diabetic Hispanic, overweight, not on insulin
- Patient 7-9: Nondiabetic Caucasian and obese (BMI>30)
- Patient 10: Diabetic Hispanic, overweight, on insulin
- Patient 11: Nondiabetic African American and overweight
- Patient 12: Diabetic Hispanic on insulin

20 Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999:

25 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose

Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated

Donor 2 and 3 AD: Adipose, Adipose Differentiated

30 Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all

cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

GO Adipose = Greater Omentum Adipose
SK = Skeletal Muscle
UT = Uterus
PL = Placenta
AD = Adipose Differentiated
AM = Adipose Midway Differentiated
U = Undifferentiated Stem Cells

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus pallidus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus pallidus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were

examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

- 5 PSP = Progressive supranuclear palsy
- Sub Nigra = Substantia nigra
- Glob Palladus= Globus palladus
- Temp Pole = Temporal pole
- Cing Gyr = Cingulate gyrus
- 10 BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and
15 Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from
20 Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD
25 senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus;
30 the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

Ctx		Occipital Ctx	
AD 6 Inf Temporal Ctx	18.0	Control (Path) 4 Occipital Ctx	13.0
AD 6 Sup Temporal Ctx	28.7	Control 1 Parietal Ctx	1.0
Control 1 Temporal Ctx	1.0	Control 2 Parietal Ctx	26.6
Control 2 Temporal Ctx	31.6	Control 3 Parietal Ctx	18.2
Control 3 Temporal Ctx	9.0	Control (Path) 1 Parietal Ctx	71.2
Control 3 Temporal Ctx	2.2	Control (Path) 2 Parietal Ctx	17.7
Control (Path) 1 Temporal Ctx	52.5	Control (Path) 3 Parietal Ctx	0.6
Control (Path) 2 Temporal Ctx	32.1	Control (Path) 4 Parietal Ctx	44.1

Table 12. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag2503, Run 208015585	Rel. Exp.(%) Ag2503, Run 212142287	Tissue Name	Rel. Exp.(%) Ag2503, Run 208015585	Rel. Exp.(%) Ag2503, Run 212142287
Adipose	0.0	0.0	Renal ca. TK-10	0.0	0.0
Melanoma* Hs688(A).T	0.0	0.0	Bladder	0.1	0.1
Melanoma* Hs688(B).T	0.0	0.0	Gastric ca. (liver met.) NCI-N87	0.1	0.1
Melanoma* M14	0.0	0.0	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.1	Colon ca. SW-948	0.0	0.0
Melanoma* SK- MEL-5	0.0	0.1	Colon ca. SW480	0.0	0.1
Squamous cell carcinoma SCC- 4	0.0	0.0	Colon ca.* (SW480 met) SW620	0.0	0.0
Testis Pool	0.2	0.3	Colon ca. HT29	0.1	0.1
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	6.4	7.8	Colon ca. CaCo-2	0.0	0.0
Placenta	0.0	0.0	Colon cancer tissue	0.1	0.2
Uterus Pool	0.0	0.0	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	0.0	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK- OV-3	0.0	0.1	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.0	0.0	Colon Pool	0.2	0.1
Ovarian ca. OVCAR-5	8.4	7.2	Small Intestine Pool	0.2	0.4
Ovarian ca. IGROV-1	0.0	0.0	Stomach Pool	0.2	0.0
Ovarian ca. OVCAR-8	0.0	0.0	Bone Marrow Pool	0.0	0.0
Ovary	0.0	0.1	Fetal Heart	0.0	0.0

(Fetal)			N		
Skeletal muscle	0.0	0.0	Ovary	0.0	0.0
Bone marrow	0.0	0.0	Ovarian ca. OVCAR-3	0.0	0.0
Thymus	0.0	0.0	Ovarian ca. OVCAR-4	0.0	0.0
Spleen	0.0	0.5	Ovarian ca. OVCAR-5	1.5	1.0
Lymph node	0.0	0.0	Ovarian ca. OVCAR-8	0.0	0.0
Colorectal	0.1	0.0	Ovarian ca. IGROV-1	0.0	0.0
Stomach	0.0	0.0	Ovarian ca. (ascites) SK-OV-3	0.0	0.3
Small intestine	0.1	0.2	Uterus	0.2	0.0
Colon ca. SW480	0.0	0.0	Placenta	0.0	0.0
Colon ca.* SW620 (SW480 met)	0.0	0.0	Prostate	3.2	1.7
Colon ca. HT29	0.1	0.0	Prostate ca.* (bone met) PC-3	0.0	0.0
Colon ca. HCT-116	0.0	0.0	Testis	0.0	0.0
Colon ca. CaCo-2	0.0	0.0	Melanoma Hs688(A).T	0.0	0.0
CC Well to Mod Diff (ODO3866)	0.1	0.2	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC-2998	0.0	0.1	Melanoma UACC-62	0.0	0.0
Gastric ca. (liver met) NCI-N87	0.0	0.1	Melanoma M14	0.0	0.0
Bladder	0.0	0.0	Melanoma LOX IMVI	0.0	0.0
Trachea	0.1	0.0	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney	0.0	0.0	Adipose	0.0	0.0

Table 14. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2503, Run 160838287	Rel. Exp.(%) Ag2503, Run 164993346	Tissue Name	Rel. Exp.(%) Ag2503, Run 160838287	Rel. Exp.(%) Ag2503, Run 164993346
Normal Colon	2.7	2.3	Kidney Margin 8120608	0.0	0.0
CC Well to Mod Diff (ODO3866)	0.1	0.2	Kidney Cancer 8120613	0.0	0.0
CC Margin (ODO3866)	0.1	0.3	Kidney Margin 8120614	0.0	0.1
CC Gr.2 rectosigmoid (ODO3868)	0.1	0.0	Kidney Cancer 9010320	0.0	0.1
CC Margin (ODO3868)	0.0	0.1	Kidney Margin 9010321	0.0	0.0
CC Mod Diff	0.1	0.1	Normal Uterus	0.0	0.0

adenocarcinoma			carcinma (transitional cell)		
SW-948- Colon adenocarcinoma	0.0	0.0	A204- Rhabdomyosarcoma	0.0	0.0
SW-480- Colon adenocarcinoma	0.0	0.0	HT-1080- Fibrosarcoma	0.0	0.0
NCI-SNU-5- Gastric carcinoma	0.0	0.0	MG-63- Osteosarcoma	0.0	0.0
KATO III- Gastric carcinoma	5.5	15.3	SK-LMS-1- Leiomyosarcoma (vulva)	1.3	0.0
NCI-SNU-16- Gastric carcinoma	0.0	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	2.5	0.0
NCI-SNU-1- Gastric carcinoma	0.0	0.0	A431- Epidermoid carcinoma	0.0	0.0
RF-1- Gastric adenocarcinoma	0.0	0.0	WM266-4- Melanoma	1.6	0.0
RF-48- Gastric adenocarcinoma	0.0	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0	0.0
MKN-45- Gastric carcinoma	0.0	0.0	MDA-MB-468- Breast adenocarcinoma	0.0	0.0
NCI-N87- Gastric carcinoma	0.0	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0	0.0
OVCAR-5- Ovarian carcinoma	3.1	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0	0.0
RL95-2- Uterine carcinoma	0.0	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0	0.0
HelaS3- Cervical adenocarcinoma	0.0	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0	0.0

Table 16. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag2503, Run 171656392	Tissue Name	Rel. Exp.(%) Ag2503, Run 171656392
BA4 Control	31.9	BA17 PSP	22.2
BA4 Control2	65.1	BA17 PSP2	10.1
BA4 Alzheimer's2	4.0	Sub Nigra Control	14.8
BA4 Parkinson's	70.2	Sub Nigra Control2	9.9
BA4 Parkinson's2	100.0	Sub Nigra Alzheimer's2	2.9
BA4 Huntington's	41.2	Sub Nigra Parkinson's2	18.9
BA4 Huntington's2	2.1	Sub Nigra Huntington's	10.5
BA4 PSP	5.3	Sub Nigra Huntington's2	10.7
BA4 PSP2	24.7	Sub Nigra PSP2	0.9
BA4 Depression	10.8	Sub Nigra Depression	0.5
BA4 Depression2	9.0	Sub Nigra Depression2	3.1
BA7 Control	42.3	Glob Palladus Control	0.5
BA7 Control2	40.9	Glob Palladus Control2	1.4
BA7 Alzheimer's2	6.4	Glob Palladus Alzheimer's	3.1
BA7 Parkinson's	18.8	Glob Palladus Alzheimer's2	1.0

agents that modulate the function of this gene product could potentially reduce neuronal degeneration in patients with Alzheimer's Disease and other neurodegenerative diseases.

In addition, defective potassium channels are known to cause several CNS disorders, including epilepsy and episodic ataxia with myokymia. Therefore, modulation of the expression or function of this gene product may potentially be useful as a treatment for the symptoms produced by ataxia and epilepsy (Jhamandas et al., Cellular Mechanisms for Amyloid beta-Protein Activation of Rat Cholinergic Basal Forebrain Neurons. *J Neurophysiol* 86(3):1312-20, 2001; Chi et al., Potassium channel openers prevent beta-amyloid toxicity in bovine vascular endothelial cells. *Neurosci Lett* 290(1):9-12, 2000; Piccini et al., Endogenous APP derivatives oppositely modulate apoptosis through an autocrine loop. *Neuroreport* 11(7):1375-9, 2000; Yu et al., Enhancement of outward potassium current may participate in beta-amyloid peptide-induced cortical neuronal death. *Neurobiol Dis* 5(2):81-8, 1998; Colom et al., Role of potassium channels in amyloid-induced cell death. *J Neurochem* 70(5):1925-34, 1998).

General_screening_panel_v1.4 Summary: Ag2503 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression in the brain. Please see *CNS_neurodegeneration_v1.0* for discussion of potential utility in the central nervous system.

There is also moderate to low expression in normal prostate and in cell lines derived from breast, lung, and ovarian cancer. Thus, this expression could be used as a diagnostic marker for the presence of cancers in any of those tissues. Furthermore, inhibition of the activity of the gene product by antibodies or small molecule inhibitors could potentially be used as a treatment of these cancers.

In both experiments, there are also significantly higher levels of expression in the fetal kidney (CTs=30-31) when compared to the adult kidney (CTs=35-36). Thus, expression of this gene could be used to differentiate between adult and fetal sources of this tissue. Furthermore, the higher levels of expression in the fetal kidney suggest that this gene product may be involved in the development of this organ. Therefore, therapeutic modulation of the expression or function of the protein encoded by this gene may be useful in the treatment of diseases of the kidney.

Among tissues with metabolic function, the expression of this potassium channel homolog is highest in the pituitary gland and shows very good concordance between the two independent runs. Potassium channels are involved in regulation of secretion in pituitary cells

and their modulation by therapeutics such as small molecule inhibitors or antibodies could be used to modulate specific secretory activities in the pituitary.

Panel 1.3D Summary: Ag2503 Two experiments with the same probe and primer set produce results that are in very good agreement, with highest expression in both experiments seen in the brain. Please see CNS_neurodegeneration_v1.0 for discussion of potential utility in the central nervous system.

Moderate to low expression is also observed in some cancer cell lines (lung and ovary) as well as normal prostate and breast. Thus, this expression could be used as a diagnostic marker for lung and ovarian cancers. Furthermore, inhibition of the activity of this gene product through the application of antibodies or small molecule inhibitors could effective in the treatment of lung or ovarian cancers.

As in panel 1.4, expression of the NOV1 gene among metabolic tissues is highest in the pituitary. Significantly lower levels of expression are seen in the adrenal gland and in fetal skeletal muscle. Potassium channels are involved in regulation of secretion in pituitary cells and their modulation by therapeutics such as small molecule inhibitors or antibodies could be used to modulate specific secretory activities in the pituitary, as well as in other tissues.

In both experiments, there is also significantly higher levels of expression in fetal skeletal muscle (CTs=33) when compared to expression in adult skeletal muscle (CTs=40). Thus, expression of the NOV1 gene could be used to differentiate between adult and fetal sources of this tissue. Furthermore, the higher levels of expression in fetal skeletal muscle suggest that this gene product may be involved in the development of the skeletal muscle in the fetus. Therefore, therapeutic modulation of the expression or function of the protein encoded by the NOV1 gene may be useful in the adult to restore mass or function to weak or dystrophic muscle.

Panel 2D Summary: Ag2503 The expression of the NOV1 gene shows good concordance between two independent runs, with highest expression in a breast cancer sample (CTs= 25-27). The expression of the NOV1 gene is increased in breast and prostate cancer compared to the normal adjacent tissue. Thus, expression of the NOV1 gene could be used as a diagnostic marker for the presence of breast and prostate cancers. Furthermore, therapeutic inhibition of the activity of the NOV1 gene through the application of antibodies or small molecule inhibitors could be effective in the treatment of these cancers.

Panel 3D Summary: Ag2503 The expression of the NOV1 gene shows good concordance between two independent runs. The highest level of expression is seen in a lung cancer cell line (NCI-H146) (CTs=30-33) . Thus, the expression of the NOV1 gene could

potentially be used as a diagnostic marker for lung cancer. Furthermore, inhibition of the activity of the protein encoded by the NOV1 gene may also be useful in the treatment of lung cancer.

5 **Panel 4D Summary** Ag2503 Data from one experiment with this probe and primer set is not included. A bad amp plot indicates that there were experimental difficulties with this run.

Panel CNS_1 Summary: Ag2503 Ubiquitous expression in this panel confirms the presence in the brain of this protein product. Please see CNS_neurodegeneration_v1.0 for discussion of potential utility in the central nervous system.

10 B. NOV2: Galanin Receptor Type 1 (GALR1)-like

Expression of the NOV2 gene (CG50293-01) was assessed using the primer-probe set Ag2534, described in Table 17. Results of the RTQ-PCR runs are shown in Tables 18, and 19. Table 17. Probe Name Ag2534

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-catagccctgtctcaagtcttg-3'	22	858	106
Probe	TET-5'-ttccatctcttcagcaaatctctca-3'-TAMRA	26	885	107
Reverse	5'-actcttcgacatcacaagaaa-3'	22	913	108

Table 18. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2534, Run 165531311	Tissue Name	Rel. Exp.(%) Ag2534, Run 165531311
Liver adenocarcinoma	5.3	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	3.0
Pancreatic ca. CAPAN 2	4.2	Renal ca. A498	7.2
Adrenal gland	0.0	Renal ca. RFX 393	2.2
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	3.7
Pituitary gland	0.0	Renal ca. TK-10	1.4
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	6.6	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	4.4
Brain (cerebellum)	1.7	Lung	0.0
Brain (hippocampus)	4.3	Lung (fetal)	0.0
Brain (substantia nigra)	3.7	Lung ca. (small cell) LX-1	3.3
Brain (thalamus)	8.7	Lung ca. (small cell) NCI-H69	5.2
Cerebral Cortex	10.5	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	2.2	Lung ca. (large cell)NCI-H460	10.6

Secondary Tr1 act	25.0	HUVEC TNF alpha + IFN gamma	5.1
Secondary Th1 rest	1.4	HUVEC TNF alpha + IL4	11.0
Secondary Th2 rest	5.6	HUVEC IL-11	11.9
Secondary Tr1 rest	2.7	Lung Microvascular EC none	7.1
Primary Th1 act	36.9	Lung Microvascular EC TNFalpha + IL-1beta	34.2
Primary Th2 act	35.6	Microvascular Dermal EC none	8.7
Primary Tr1 act	48.3	Microvascular Dermal EC TNFalpha + IL-1beta	5.7
Primary Th1 rest	35.6	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	21.0	Small airway epithelium none	13.9
Primary Tr1 rest	14.1	Small airway epithelium TNFalpha + IL-1beta	12.0
CD45RA CD4 lymphocyte act	8.0	Coronary artery SMC rest	7.5
CD45RO CD4 lymphocyte act	25.0	Coronary artery SMC TNFalpha + IL-1beta	2.9
CD8 lymphocyte act	21.6	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	19.9	Astrocytes TNFalpha + IL-1beta	5.2
Secondary CD8 lymphocyte act	5.7	KU-812 (Basophil) rest	17.3
CD4 lymphocyte none	4.6	KU-812 (Basophil) PMA/ionomycin	24.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	11.7	CCD1106 (Keratinocytes) none	14.3
LAK cells rest	11.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	5.5
LAK cells IL-2	22.7	Liver cirrhosis	6.1
LAK cells IL-2+IL-12	22.1	Lupus kidney	3.3
LAK cells IL-2+IFN gamma	26.2	NCI-H292 none	50.0
LAK cells IL-2+ IL-18	23.0	NCI-H292 IL-4	46.7
LAK cells PMA/ionomycin	6.1	NCI-H292 IL-9	24.7
NK Cells IL-2 rest	18.0	NCI-H292 IL-13	11.7
Two Way MLR 3 day	12.5	NCI-H292 IFN gamma	19.2
Two Way MLR 5 day	13.7	HPAEC none	3.0
Two Way MLR 7 day	6.4	HPAEC TNF alpha + IL-1 beta	10.6
PBMC rest	4.0	Lung fibroblast none	5.8
PBMC PWM	47.6	Lung fibroblast TNF alpha + IL-1 beta	3.0
PBMC PHA-L	19.5	Lung fibroblast IL-4	24.0
Ramos (B cell) none	46.7	Lung fibroblast IL-9	19.8
Ramos (B cell) ionomycin	100.0	Lung fibroblast IL-13	5.3
B lymphocytes PWM	37.6	Lung fibroblast IFN gamma	11.4
B lymphocytes CD40L and IL-4	30.1	Dermal fibroblast CCD1070 rest	12.9
EOL-1 dbcAMP	11.7	Dermal fibroblast CCD1070	41.8

B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	1.3	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	9.6
Macrophages rest	0.0	Lung	5.6
Macrophages LPS	0.0	Thymus	100.0
HUVEC none	0.0	Kidney	0.6
HUVEC starved	0.0		

Panel 1.3D Summary: Ag1905 Two experiments with the same probe and primer set produce results that are in good agreement with highest expression in the lung cancer cell line SHP-77 (CTs=30) and the trachea (CTs=30-31). There is also significant expression of the NOV3 gene in cell lines derived from the colon and ovary. This gene may play a role in different types of lung, ovary and colon cancer as it is more highly expressed in cell lines derived from these cancers compared to the normal tissues. Furthermore, expression in normal brain and pancreas seems to be higher than cancer cell lines derived from these tissues. Thus, expression of the NOV3 gene could be used as a marker or as a therapeutic for colon, ovarian, brain, lung, and pancreatic cancer. In addition, therapeutic modulation of the product of this gene, through the use of peptides, chimeric molecules or small molecule drugs, may be useful in the therapy of these cancers.

There is also significant expression of the NOV3 gene in tissues involved in the central nervous system including the amygdala, hippocampus, thalamus, cerebral cortex, and spinal cord.

Purinoreceptors found in GDNF sensitive sensory neurons mediate nociceptor function. Since the NOV3 gene product is a homolog of a purinoreceptor, agents that block the action of this receptor may have utility in treating pain, either acting as analgesics or inhibiting the establishment of chronic pain. In addition, since adenosine plays a significant neuromodulatory role in brain regions such as the hippocampus, cortex, basal ganglia, and thalamus, the NOV3 purinoreceptor-homolog is localized in a position to participate with the action of adenosine in these brain regions. The protein encoded by the NOV3 gene is most homologous to P2Y4 and P2Y6 purinoreceptors, suggesting that its function may be similar to the PLC-mediated Ca²⁺ mobilization induced by these receptors. Ca²⁺ mobilization is an

important component of the molecular process leading to neurotransmitter release. Adenosine modulates the release of glutamate in the brain, which is the main excitatory amino acid neurotransmitter. Glutamate exerts excitotoxic neuronal damage and death in a number of pathological conditions, including stroke. Agonists of A1 adenosine receptors attenuate this damage via G protein-coupled inhibition of glutamate release. Antagonists of A2 receptors also attenuate glutamate induced excitotoxicity. Therefore, agents that inhibit or stimulate the protein encoded by the NOV3 gene are likely to affect glutamate release in the brain and the subsequent action of glutamate in these regions. If the NOV3 gene product functions similarly to the A1 receptor with respect to glutamate release, then agonists of the putative receptor are likely to have utility in the treatment of stroke. If the NOV3 gene product functions similarly to the A2 receptor, then antagonists of the putative receptor are likely to have utility in the treatment of stroke. Furthermore, antagonists of the A2a purinoceptor are antidepressants. Therefore, antagonists of the NOV3 gene product may be useful antidepressants. A2a receptor antagonists also counter parkinsonian-like symptoms in mice, suggesting that the NOV3 gene product antagonists may also have utility in the treatment of Parkinson's disease (Liu et al., P2Y purinoceptor activation mobilizes intracellular Ca²⁺ and induces a membrane current in rat intracardiac neurones. *J Physiol.* 526 Pt 2:287-98, 2000; Ongini et al., Selective adenosine A2A receptor antagonists. *Farmacol.* 56(1-2):87-90, 2001; Chen et al., Neuroprotection by caffeine and A(2A) adenosine receptor inactivation in a model of Parkinson's disease. *J Neurosci.* 21:RC143, 2001; Wardas et al., SCH 58261, an A(2A) adenosine receptor antagonist, counteracts parkinsonian-like muscle rigidity in rats. *Synapse.* 41:160-71, 2001; Driessen et al., Depression of C fiber-evoked activity by intrathecally administered reactive red 2 in rat thalamic neurons. *Brain Res.* 796 (12):284-90, 1998; El Yacoubi et al., Adenosine A2A receptor antagonists are potential antidepressants: evidence based on pharmacology and A2A receptor knockout mice. *Br J Pharmacol.* 134:68-77, 2001).

Panel 2D Summary: Ag1905 Highest expression of the NOV3 gene is detected in a colon cancer (CT=30.4). Furthermore, expression of this gene appears to be overexpressed in colon cancer when compared to normal adjacent tissue in all six matched tissue pairs present in this panel. Thus, expression of the NOV3 gene could be used to differentiate between colon cancer and normal tissue. Furthermore, therapeutic modulation of the function or activity of the NOV3 gene product could be effective in the treatment of colon cancer. The NOV3 gene also shows a reverse association in the kidney, with overexpression of the gene present in normal kidney when compared to the corresponding cancerous tissue. Thus, expression of the

NK Cells IL-2 rest	1.4	NCI-H292 IL-13	23.5
Two Way MLR 3 day	2.5	NCI-H292 IFN gamma	26.4
Two Way MLR 5 day	1.0	HPAEC none	1.0
Two Way MLR 7 day	1.3	HPAEC TNF alpha + IL-1 beta	0.5
PBMC rest	0.9	Lung fibroblast none	17.1
PBMC PWM	7.7	Lung fibroblast TNF alpha + IL-1 beta	5.9
PBMC PHA-L	4.3	Lung fibroblast IL-4	46.0
Ramos (B cell) none	10.9	Lung fibroblast IL-9	53.6
Ramos (B cell) ionomycin	32.5	Lung fibroblast IL-13	34.9
B lymphocytes PWM	7.5	Lung fibroblast IFN gamma	65.1
B lymphocytes CD40L and IL-4	8.4	Dermal fibroblast CCD1070 rest	100.0
EOL-1 dbcAMP	0.6	Dermal fibroblast CCD1070 TNF alpha	66.9
EOL-1 dbcAMP PMA/ionomycin	0.3	Dermal fibroblast CCD1070 IL-1 beta	24.0
Dendritic cells none	0.2	Dermal fibroblast IFN gamma	26.4
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	21.3
Dendritic cells anti-CD40	0.2	IBD Colitis 2	0.3
Monocytes rest	0.2	IBD Crohn's	2.2
Monocytes LPS	0.2	Colon	26.4
Macrophages rest	0.6	Lung	29.9
Macrophages LPS	0.2	Thymus	34.4
HUVEC none	2.7	Kidney	17.4
HUVEC starved	6.7		

CNS_neurodegeneration_v1.0 Summary: Ag2510 This panel does not show any evident association between expression levels and disease in the conditions examined.

However, these results confirm expression of the NOV4 gene in the brain. Please see Panel 1.3D for discussion of potential utility in the central nervous system.

Panel 1.3D Summary: Ag2510 The NOV4 gene is expressed at a low to moderate levels in most of the tissues and cell lines on this panel, with highest expression is seen in a lung cancer cell line (CT=28.62). This ubiquitous expression suggests it plays a role in cell survival and proliferation for a majority of cell types. There appears to be lower expression in normal prostate, pancreas and ovary compared to the tumor cell lines from these tissues. Hence, expression of the NOV4 gene could potentially be used as a diagnostic marker for these cancers.

Among metabolic tissues, the expression of this LOMP-like gene is strongest in the heart and skeletal muscle, while fetal heart and fetal skeletal muscle express very low levels of this gene. Therefore, the NOV4 gene could be used to differentiate between the adult and fetal states of these two tissues. Lower levels of the NOV4 gene are expressed in the stomach,

pancreas, salivary gland, small intestine, liver, fetal liver and adipose. The LOMP-like protein encoded by the NOV4 gene may be involved in protein-protein interactions within the cell as evidenced by the presence of the LIM and PDZ domains within its sequence. Therefore, modulating expression or activity of the NOV4 gene may affect the development or physiological activity of the organs expressing this gene.

Moderate expression throughout the brain and specifically in the amygdala, hippocampus, substantia nigra, cortex and cerebellum indicate a possible function in these CNS tissues.

LIM domain only proteins contain only the LIM domain that mediates protein-protein interaction, while other domains, such as zinc finger domains, are not present. Thus, these proteins may act as inhibitors of LIM proteins with functions dependent on the absent domains. LIM domain proteins play a role in neuronal and pituitary development. Therefore, agents that modulate the expression or function of the protein encoded by the NOV4 gene may be useful in diseases of dysregulated neural development, such as autism or ataxia (Putilina et al. Analysis of a human cDNA containing a tissue-specific alternatively spliced LIM domain. Biochem Biophys Res Commun 252(2):433-9, 1998; Netchine et al., Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. Nat Genet 25(2):182-6, 2000).

Panel 2.2 Summary: Ag2510 The NOV4 gene is expressed at low to moderate levels in most of the tissues on this panel, with highest expression in a lung margin sample (CT=28.4). Expression of the NOV4 gene appears to be overexpressed in normal lung when compared to cancer in 5 of the 7 tissue pairs present in this panel, as well as decreased expression in a metastatic cancer to the liver compared to the normal adjacent tissue. Thus, decreased expression of the NOV4 protein could potentially be used as a diagnostic marker for the presence of these cancers. Furthermore, increasing the activity or expression of the NOV4 gene product could act as a therapy for lung cancer.

Panel 4D Summary: Ag2510 This transcript is highly expressed in fibroblasts and endothelial cells regardless of treatment, with highest expression in dermal fibroblasts (CT=26). This transcript is also highly expressed in normal tissue. Thus, this expression profile suggests that this transcript or the NOV4 protein it encodes could be used to identify fibroblasts and endothelial cells.

E. NOV5: Epidermal Growth Factor-like

Temporal Ctx			1 Parietal Ctx		
Control 3 Temporal Ctx	6.5	14.6	Control (Path) 2 Parietal Ctx	15.7	20.2
Control (Path) 1 Temporal Ctx	18.0	21.6	Control (Path) 3 Parietal Ctx	2.6	4.0
Control (Path) 2 Temporal Ctx	13.9	22.1	Control (Path) 4 Parietal Ctx	21.9	25.7

Table 31. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2505, Run 165531061	Tissue Name	Rel. Exp.(%) Ag2505, Run 165531061
Liver adenocarcinoma	1.8	Kidney (fetal)	27.5
Pancreas	13.7	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	1.5	Renal ca. A498	0.2
Adrenal gland	3.6	Renal ca. RXF 393	39.8
Thyroid	100.0	Renal ca. ACHN	51.1
Salivary gland	4.1	Renal ca. UO-31	0.2
Pituitary gland	37.6	Renal ca. TK-10	0.0
Brain (fetal)	44.1	Liver	1.4
Brain (whole)	9.3	Liver (fetal)	2.3
Brain (amygdala)	8.1	Liver ca. (hepatoblast) HepG2	11.8
Brain (cerebellum)	1.8	Lung	75.3
Brain (hippocampus)	10.2	Lung (fetal)	54.7
Brain (substantia nigra)	29.3	Lung ca. (small cell) LX- 1	5.5
Brain (thalamus)	3.6	Lung ca. (small cell) NCI-H69	5.6
Cerebral Cortex	7.7	Lung ca. (s.cell var.) SHP-77	0.2
Spinal cord	15.2	Lung ca. (large cell) NCI- H460	1.2
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	1.2
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	3.2
astrocytoma SW1783	0.3	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.4	Lung ca. (non-s.cl) NCI- H522	0.0
astrocytoma SF-539	1.8	Lung ca. (squam.) SW 900	1.8
astrocytoma SNB-75	2.7	Lung ca. (squam.) NCI- H596	14.6
glioma SNB-19	0.0	Mammary gland	11.9
glioma U251	9.3	Breast ca.* (pl.ef) MCF- 7	89.5
glioma SF-295	0.4	Breast ca.* (pl.ef) MDA- MB-231	0.0
Heart (Fetal)	10.0	Breast ca.* (pl. ef) T47D	24.7

Heart	3.1	Breast ca. BT-549	2.3
Skeletal muscle (Fetal)	12.8	Breast ca. MDA-N	0.0
Skeletal muscle	20.9	Ovary	3.5
Bone marrow	1.2	Ovarian ca. OVCAR-3	6.4
Thymus	6.0	Ovarian ca. OVCAR-4	0.0
Spleen	6.7	Ovarian ca. OVCAR-5	0.0
Lymph node	6.7	Ovarian ca. OVCAR-8	0.2
Colorectal	23.5	Ovarian ca. IGROV-1	14.5
Stomach	12.0	Ovarian ca. (ascites) SK-OV-3	9.3
Small intestine	54.3	Uterus	27.7
Colon ca. SW480	1.1	Placenta	2.9
Colon ca. * SW620 (SW480 met)	1.4	Prostate	25.0
Colon ca. HT29	7.3	Prostate ca. * (bone met) PC-3	0.0
Colon ca. HCT-116	7.3	Testis	2.5
Colon ca. CaCo-2	10.7	Melanoma Hs688(A).T	0.0
CC Well to Mod Diff (ODO3866)	8.5	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	2.9	Melanoma UACC-62	0.0
Gastric ca. (liver met) NCI-N87	71.7	Melanoma M14	0.0
Bladder	14.8	Melanoma LOX IMVI	0.0
Trachea	21.9	Melanoma* (met) SK-MEL-5	0.0
Kidney	38.2	Adipose	19.3

Table 32. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2505, Run 164318134	Tissue Name	Rel. Exp.(%) Ag2505, Run 164318134
Secondary Th1 act	0.0	HUVEC IL-1beta	0.2
Secondary Th2 act	0.0	HUVEC IFN gamma	0.1
Secondary Tr1 act	0.2	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.1
Secondary Th2 rest	0.3	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.2	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.1	Microvascular Dermal EC TNFalpha + IL-1beta	0.2
Primary Th1 rest	0.1	Bronchial epithelium TNFalpha + IL1beta	2.0
Primary Th2 rest	0.0	Small airway epithelium none	0.5
Primary Tr1 rest	0.1	Small airway epithelium TNFalpha + IL-1beta	19.6
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte	0.0	Coronary artery SMC TNFalpha	0.0

[REDACTED]

believed to be a beneficial processing alternative to beta secretase, is increased by EGF in neuronal cells. This suggests that the increased expression observed here is a compensatory action in the brain to counter the mechanisms of Alzheimer's Disease. Therefore, the protein encoded by the NOV5 gene may be a potential therapeutic agent for the treatment of

5 Alzheimer's disease and other neurodegenerative diseases.

EGF is also known to facilitate long term potentiation (LTP) in the hippocampus, a process thought to underlie learning and memory. Therefore, the NOV5 gene may have utility in treating disorders of memory, such as neurodegenerative diseases and aging, when used alone or in combination with other growth factors such as bFGF.

10 In addition, EGF supports the growth and differentiation of dopaminergic neurons, which are selectively vulnerable to loss in Parkinson's disease. Therefore, the NOV5 gene product may have utility in treating Parkinson's Disease (Slack et al., Rapid stimulation of amyloid precursor protein release by epidermal growth factor: role of protein kinase C. *Biochem J* 327 (Pt 1):245-9, 1997; Abe et al., Effects of epidermal growth factor and basic

15 fibroblast growth factor on generation of long-term potentiation in the dentate gyrus of fimbria-fornix-lesioned rats. *Brain Res* 593(2):335-8, 1992; Storch et al., Long-term proliferation and dopaminergic differentiation of human mesencephalic neural precursor cells. *Exp Neurol* 170(2):317-25, 2001).

Panel 1.3D Summary: Ag2505: Highest expression of the NOV5 gene is in the

20 thyroid (CT=29.3), with lower but still significant levels of expression seen in other metabolic tissues, including skeletal muscle, fetal skeletal muscle, small intestine, stomach, pancreas, adipose and fetal heart. Very low levels are also seen in heart and placenta. The NOV5 gene encodes a putative novel adhesion molecule. Studies in mouse have revealed a gene (perhaps the mouse ortholog of this human gene) very homologous to the NOV5 gene that is called

25 POEM (preosteoblast epidermal growth factor-like repeat protein with meprin, A5 protein, and receptor protein-tyrosine phosphatase mu domain) or nephronectin. POEM/nephronectin seems to be a ligand for the alpha(8)beta(1) integrin as evidenced by two independent sets of published data. Integrins are known to mediate development and organogenesis. Other known ligands for the alpha(8)beta(1) integrin include fibronectin, vitronectin, tenascin, and

30 osteopontin. Therefore, modulation of the expression or activity of the NOV5 gene product by protein or antibody therapeutics may be an effective therapeutic for disorders involving alpha(8)beta(1) integrin signaling.

Overall, the NOV5 gene is expressed at a low to moderate level in the normal tissues on this panel. Furthermore, the brain, prostate, lung and colon cancer cell lines show a very

low level of expression compared to the normal organs. This suggests that this molecule can potentially be used as a therapeutic inhibitor for these cancers.

Expression in the brain in the substantia nigra, hippocampus, cortex, amygdala, thalamus and spinal cord indicates a further functional role for the NOV5 gene product in CNS processes mediated by these regions. Please see CNS_neurodegeneration_v1.0 for discussion of utility in the central nervous system (Morimura et al., Molecular cloning of POEM: a novel adhesion molecule that interacts with alpha8beta1 integrin. J Biol Chem 276(45):42172-81, 2001; Brandenberger et al., Identification and characterization of a novel extracellular matrix protein nephronectin that is associated with integrin alpha8beta1 in the embryonic kidney. J Cell Biol 154(2):447-58, 2001; Schwartz et al., Integrins: emerging paradigms of signal transduction. Annu. Rev. Cell Dev. Biol. 11, 549-599, 1995; Clark and Brugge, Integrins and signal transduction pathways: the road taken. Science 268, 233-239, 1995).

Panel 4D Summary: Ag 2505 Highest expression of this transcript is found in the thymus and the lung (CTs=27-28). Consistent with this lung expression, this transcript is found in the pulmonary mucoepidermoid cell line H292 and is up-regulated upon treatment with the Th2 cytokines IL4 and IL9. The NOV5 gene is also expressed at lower levels in lung fibroblasts treated with IL4. This transcript profile suggests that modulation of the expression or activity of the NOV5 gene product by protein or antibody therapeutics may be beneficial for the treatment of inflammatory lung diseases such as asthma, emphysema and chronic obstructive pulmonary diseases. Furthermore, therapeutics designed with the protein encoded for by this transcript could be important for maintaining or restoring normal function of thymus during inflammation.

F. NOV6a – NOV6c: Hyaluronan Mediated Motility Receptor-like

Expression of the NOV6a gene (CG50239-01) and variants NOV6b (CG50239-02) and NOV6c (CG50239-03) was assessed using the primer-probe set Ag1901, described in Table 33. Results of the RTQ-PCR runs are shown in Tables 34-37.

Table 33. Probe Name Ag1901

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggaggctgaactggagaaaa-3'	20	1260	118
Probe	TET-5'-ccacctgcttttgcaggaaaagtat-3'-TAMRA	26	1301	119
Reverse	5'-cttcaaggctttgcaccata-3'	20	1332	120

Table 34. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag1901, Run 162459326	Tissue Name	Rel. Exp.(%) Ag1901, Run 162459326
-------------	------------------------------------	-------------	------------------------------------

Liver adenocarcinoma	33.9	Kidney (fetal)	3.6
Pancreas	0.0	Renal ca. 786-0	17.1
Pancreatic ca. CAPAN 2	22.2	Renal ca. A498	11.2
Adrenal gland	0.5	Renal ca. RXF 393	4.8
Thyroid	0.5	Renal ca. ACHN	2.5
Salivary gland	0.1	Renal ca. UO-31	28.1
Pituitary gland	0.0	Renal ca. TK-10	7.4
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	13.8
Brain (amygdala)	0.2	Liver ca. (hepatoblast) HepG2	9.6
Brain (cerebellum)	0.0	Lung	0.3
Brain (hippocampus)	0.3	Lung (fetal)	1.8
Brain (substantia nigra)	0.2	Lung ca. (small cell) LX-1	10.8
Brain (thalamus)	0.6	Lung ca. (small cell) NCI-H69	9.7
Cerebral Cortex	0.2	Lung ca. (s.cell var.) SHP-77	69.3
Spinal cord	0.8	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	18.2	Lung ca. (non-sm. cell) A549	1.0
glio/astro U-118-MG	14.9	Lung ca. (non-s.cell) NCI-H23	25.2
astrocytoma SW1783	100.0	Lung ca. (non-s.cell) HOP-62	3.4
neuro*; met SK-N-AS	8.7	Lung ca. (non-s.cl) NCI-H522	10.8
astrocytoma SF-539	17.9	Lung ca. (squam.) SW 900	10.0
astrocytoma SNB-75	9.7	Lung ca. (squam.) NCI-H596	18.6
glioma SNB-19	4.5	Mammary gland	1.6
glioma U251	3.7	Breast ca.* (pl.ef) MCF-7	35.6
glioma SF-295	1.4	Breast ca.* (pl.ef) MDA-MB-231	4.5
Heart (Fetal)	0.8	Breast ca.* (pl. ef) T47D	2.1
Heart	0.2	Breast ca. BT-549	16.3
Skeletal muscle (Fetal)	3.7	Breast ca. MDA-N	30.6
Skeletal muscle	0.5	Ovary	1.3
Bone marrow	9.5	Ovarian ca. OVCAR-3	10.5
Thymus	44.1	Ovarian ca. OVCAR-4	1.1
Spleen	1.0	Ovarian ca. OVCAR-5	19.2
Lymph node	0.5	Ovarian ca. OVCAR-8	10.1
Colorectal	1.6	Ovarian ca. IGROV-1	1.9
Stomach	2.0	Ovarian ca. (ascites) SK-OV-3	24.5
Small intestine	4.7	Uterus	0.1
Colon ca. SW480	31.0	Placenta	2.3

Colon ca.* SW620 (SW480 met)	18.9	Prostate	0.0
Colon ca. HT29	32.5	Prostate ca.* (bone met) PC-3	9.9
Colon ca. HCT-116	37.9	Testis	32.8
Colon ca. CaCo-2	40.1	Melanoma Hs688(A).T	0.8
CC Well to Mod Diff (ODO3866)	5.7	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	14.8	Melanoma UACC-62	0.0
Gastric ca. (liver met) NCI-N87	1.6	Melanoma M14	2.6
Bladder	7.2	Melanoma LOX IMVI	0.7
Trachea	1.7	Melanoma* (met) SK- MEL-5	11.7
Kidney	0.2	Adipose	1.1

Table 35. Panel 2D

Tissue Name	Rel. Exp.(%) Ag1901, Run 170858349	Tissue Name	Rel. Exp.(%) Ag1901, Run 170858349
Normal Colon	27.2	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	15.7	Kidney Cancer 8120613	0.2
CC Margin (ODO3866)	9.3	Kidney Margin 8120614	0.0
CC Gr.2 recto&sigmoid (ODO3868)	9.7	Kidney Cancer 9010320	5.6
CC Margin (ODO3868)	2.0	Kidney Margin 9010321	0.0
CC Mod Diff (ODO3920)	39.5	Normal Uterus	0.2
CC Margin (ODO3920)	7.3	Uterine Cancer 064011	3.3
CC Gr.2 ascend colon (ODO3921)	42.0	Normal Thyroid	0.3
CC Margin (ODO3921)	9.3	Thyroid Cancer	3.5
CC from Partial Hepatectomy (ODO4309) Mets	38.4	Thyroid Cancer A302152	1.6
Liver Margin (ODO4309)	1.4	Thyroid Margin A302153	1.2
Colon mets to lung (OD04451-01)	5.6	Normal Breast	3.2
Lung Margin (OD04451-02)	0.0	Breast Cancer	46.7
Normal Prostate 6546-1	4.3	Breast Cancer (OD04590-01)	37.9
Prostate Cancer (OD04410)	2.9	Breast Cancer Mets (OD04590-03)	37.4
Prostate Margin (OD04410)	6.0	Breast Cancer Metastasis	16.0
Prostate Cancer (OD04720-01)	0.8	Breast Cancer	5.5
Prostate Margin (OD04720-02)	0.6	Breast Cancer	4.5
Normal Lung	8.5	Breast Cancer 9100266	2.8
Lung Met to Muscle (ODO4286)	72.2	Breast Margin 9100265	1.1
Muscle Margin (ODO4286)	2.0	Breast Cancer A209073	20.0
Lung Malignant Cancer (OD03126)	12.9	Breast Margin A2090734	2.4
Lung Margin (OD03126)	9.5	Normal Liver	0.9

Lung Cancer (OD04404)	23.8	Liver Cancer	4.4
Lung Margin (OD04404)	1.4	Liver Cancer 1025	0.6
Lung Cancer (OD04565)	5.3	Liver Cancer 1026	0.4
Lung Margin (OD04565)	1.6	Liver Cancer 6004-T	0.2
Lung Cancer (OD04237-01)	51.4	Liver Tissue 6004-N	11.5
Lung Margin (OD04237-02)	0.7	Liver Cancer 6005-T	10.4
Ocular Mel Met to Liver (ODO4310)	2.9	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	1.2	Normal Bladder	12.7
Melanoma Metastasis	42.0	Bladder Cancer	6.0
Lung Margin (OD04321)	0.9	Bladder Cancer	17.1
Normal Kidney	2.3	Bladder Cancer (OD04718-01)	98.6
Kidney Ca, Nuclear grade 2 (OD04338)	4.2	Bladder Normal Adjacent (OD04718-03)	2.2
Kidney Margin (OD04338)	0.4	Normal Ovary	1.4
Kidney Ca Nuclear grade 1/2 (OD04339)	2.2	Ovarian Cancer	5.6
Kidney Margin (OD04339)	1.0	Ovarian Cancer (OD04768-07)	100.0
Kidney Ca, Clear cell type (OD04340)	2.4	Ovary Margin (OD04768-08)	0.9
Kidney Margin (OD04340)	0.8	Normal Stomach	5.0
Kidney Ca, Nuclear grade 3 (OD04348)	20.0	Gastric Cancer 9060358	1.0
Kidney Margin (OD04348)	2.0	Stomach Margin 9060359	1.4
Kidney Cancer (OD04622-01)	0.4	Gastric Cancer 9060395	9.9
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	3.1
Kidney Cancer (OD04450-01)	3.3	Gastric Cancer 9060397	22.2
Kidney Margin (OD04450-03)	1.2	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	0.2	Gastric Cancer 064005	47.6

Table 36. Panel 3D

Tissue Name	Rel. Exp.(%) Ag1901, Run 162460515	Tissue Name	Rel. Exp.(%) Ag1901, Run 162460515
Daoy- Medulloblastoma	7.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	28.5
TE671- Medulloblastoma	20.9	ES-2- Ovarian clear cell carcinoma	31.2
D283 Med- Medulloblastoma	27.4	Ramos- Stimulated with PMA/ionomycin 6h	52.5
PFSK-1- Primitive Neuroectodermal	31.6	Ramos- Stimulated with PMA/ionomycin 14h	23.0
XF-498- CNS	30.1	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	53.6
SNB-78- Glioma	32.5	Raji- Burkitt's lymphoma	15.4
SF-268- Glioblastoma	49.0	Daudi- Burkitt's lymphoma	58.2
T98G- Glioblastoma	61.1	U266- B-cell plasmacytoma	34.4
SK-N-SH- Neuroblastoma (metastasis)	55.1	CA46- Burkitt's lymphoma	31.0
SF-295- Glioblastoma	53.6	RL- non-Hodgkin's B-cell	15.5

		lymphoma	
Cerebellum	0.4	JM1- pre-B-cell lymphoma	30.6
Cerebellum	0.0	Jurkat- T cell leukemia	45.7
NCI-H292- Mucoepidermoid lung carcinoma	29.5	TF-1- Erythroleukemia	100.0
DMS-114- Small cell lung cancer	6.0	HUT 78- T-cell lymphoma	28.5
DMS-79- Small cell lung cancer	48.0	U937- Histiocytic lymphoma	20.2
NCI-H146- Small cell lung cancer	40.1	KU-812- Myelogenous leukemia	41.8
NCI-H526- Small cell lung cancer	35.8	769-P- Clear cell renal carcinoma	29.5
NCI-N417- Small cell lung cancer	11.7	Caki-2- Clear cell renal carcinoma	42.6
NCI-H82- Small cell lung cancer	26.4	SW 839- Clear cell renal carcinoma	16.2
NCI-H157- Squamous cell lung cancer (metastasis)	49.3	G401- Wilms' tumor	13.9
NCI-H1155- Large cell lung cancer	46.7	Hs766T- Pancreatic carcinoma (LN metastasis)	76.3
NCI-H1299- Large cell lung cancer	81.8	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	48.6
NCI-H727- Lung carcinoid	48.3	SU86.86- Pancreatic carcinoma (liver metastasis)	38.7
NCI-UMC-11- Lung carcinoid	62.9	BxPC-3- Pancreatic adenocarcinoma	16.2
LX-1- Small cell lung cancer	19.6	HPAC- Pancreatic adenocarcinoma	52.9
Colo-205- Colon cancer	23.3	MIA PaCa-2- Pancreatic carcinoma	7.1
KM12- Colon cancer	42.6	CFPAC-1- Pancreatic ductal adenocarcinoma	84.1
KM20L2- Colon cancer	10.3	PANC-1- Pancreatic epithelioid ductal carcinoma	40.1
NCI-H716- Colon cancer	31.2	T24- Bladder carcinma (transitional cell)	19.6
SW-48- Colon adenocarcinoma	8.7	5637- Bladder carcinoma	21.2
SW1116- Colon adenocarcinoma	9.0	HT-1197- Bladder carcinoma	13.3
LS 174T- Colon adenocarcinoma	41.8	UM-UC-3- Bladder carcinma (transitional cell)	23.7
SW-948- Colon adenocarcinoma	1.7	A204- Rhabdomyosarcoma	14.1
SW-480- Colon adenocarcinoma	12.2	HT-1080- Fibrosarcoma	30.6
NCI-SNU-5- Gastric carcinoma	34.2	MG-63- Osteosarcoma	24.1
KATO III- Gastric carcinoma	55.5	SK-LMS-1- Leiomyosarcoma (vulva)	52.9
NCI-SNU-16- Gastric carcinoma	8.8	SJRH30- Rhabdomyosarcoma (met to bone marrow)	18.9
NCI-SNU-1- Gastric carcinoma	29.9	A431- Epidermoid carcinoma	19.5

modulation of the NOV6 gene by peptides, small molecule drugs or antibodies may modulate the response of these tissues to specific extracellular signals or growth factors (Cheung et al., Receptor for hyaluronan-mediated motility (RHAMM), a hyaladherin that regulates cell responses to growth factors. *Biochem Soc Trans* 27(2):135-42, 1999).

5 **Panel 2D Summary:** Ag1901 The NOV6 gene is expressed at low levels in this panel, with highest expression is seen in an ovarian cancer sample (CT=31.1). The expression of the NOV6 gene is associated with colon, lung, kidney, ovary, bladder, breast and stomach cancer compared to the normal matched tissue. Thus, expression of the NOV6 gene could be used as a marker for these cancers. In addition, therapeutic modulation of the product of the
10 NOV6 gene, through the use of peptides, antibodies, chimeric molecules or small molecule drugs may be useful in the therapy of these cancers (Hall and Turley, Hyaluronan: RHAMM mediated cell locomotion and signaling in tumorigenesis. *J Neurooncol* 26(3):221-9, 1995).

Panel 3D Summary: Ag1901 Expression of the NOV6 gene is ubiquitous among all the cancer cell lines present in this panel. Please see Panel 2D for discussion of potential utility
15 of the NOV6 gene in the context of cancer.

Panel 4D Summary: Ag1901 Highest expression of the NOV6 transcript is found in activated B cells including B cells plus PWM and the activated Burkitt lymphoma cell line Ramos (CT 27.5). At a lower level, this transcript is found in activated Th1 and Tr1 cells, with expression higher than that observed in activated Th2 T cells. This transcript is encodes a
20 protein that is homologous to a receptor for a hyaluronic acid-mediated motility-like molecule. This protein has been linked to the regulation of cell locomotion and to density dependent contact inhibition of fibroblasts, smooth muscle cells, macrophages, lymphocytes, astrocytes and sperm. Modulation of the expression or activity of the encoded protein by antibodies or small molecule therapeutics could potentially be useful in preventing the recruitment of
25 activated T and B cells to inflammatory sites and may therefore be beneficial in the treatment of T and B cell mediated disease such as autoimmune diseases, rheumatoid arthritis, inflammatory bowel disease, and lupus.

 The expression of the NOV6 transcript has been reported in in vivo activated B cells such as those in Crohn's patients and in malignant B cells. Therefore, modulation of the
30 expression or activity of the encoded protein by antibodies or small molecules therapeutics may also be beneficial in the treatment of B cell malignancies or conditions that involve chronically activated B cells, including Crohn's disease and rheumatoid arthritis.

 In addition, the NOV6 gene is expressed in activated dermal fibroblasts. Therefore, modulation of the expression of the protein encoded by the NOV6 gene may also be useful for

- treatment of fibroplasias such as those associated with wound healing and other skin inflammatory diseases, such as psoriasis (Pilarski et al., RHAMM, a receptor for hyaluronan-mediated motility, on normal human lymphocytes, thymocytes and malignant B cells: a mediator in B cell malignancy? Leuk Lymphoma 14(5-6):363-74, 1994; Crainie et al.,
- 5 Overexpression of the receptor for hyaluronan-mediated motility (RHAMM) characterizes the malignant clone in multiple myeloma: identification of three distinct RHAMM variants. Blood 93(5):1684-96, 1999; Turley et al., Expression and function of a receptor for hyaluronan-mediated motility on normal and malignant B lymphocytes. Blood 81(2):446-53, 1993; Lovvorn 3rd, et al., Hyaluronan receptor expression increases in fetal excisional skin wounds
- 10 and correlates with fibroplasia. J Pediatr Surg 33(7):1062-9; discussion 1069-70, 1998).

G. NOV7: Serpin-like

Expression of the NOV7 gene (CG54308-05) was assessed using the primer-probe set Ag548, described in Table 38. Results of the RTQ-PCR runs are shown in Tables 39-42.

Table 38. Probe Name Ag548

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tggagcagctcagaaaacatgt-3'	22	789	121
Probe	TET-5'-agaatcggtgggtcctgtccttcccc-3'-TAMRA	25	815	122
Reverse	5'-catagctgtcttccagggtgaac-3'	23	842	123

15 Table 39. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag548, Run 122671023	Tissue Name	Rel. Exp.(%) Ag548, Run 122671023
Endothelial cells	0.0	Renal ca. 786-0	0.0
Heart (Fetal)	0.0	Renal ca. A498	0.0
Pancreas	0.0	Renal ca. RXF 393	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.0
Adrenal gland	0.0	Renal ca. UO-31	0.0
Thyroid	0.0	Renal ca. TK-10	0.0
Salivary gland	0.0	Liver	0.0
Pituitary gland	0.0	Liver (fetal)	0.0
Brain (fetal)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	0.0	Lung	0.0
Brain (amygdala)	0.0	Lung (fetal)	0.0
Brain (cerebellum)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (hippocampus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Brain (thalamus)	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Cerebral Cortex	0.0	Lung ca. (large cell) NCI-H460	0.0

Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	17.4	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	21.3	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	11.3
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX- I	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell) NCI- H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI- H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI- H596	14.0
glioma SNB-19	0.0	Mammary gland	23.7
glioma U251	0.0	Breast ca. * (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca. * (pl.ef) MDA- MB-231	0.0
Heart (Fetal)	0.0	Breast ca. * (pl. ef) T47D	11.5
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (Fetal)	2.1	Breast ca. MDA-N	22.8
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	100.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	9.7	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca. (ascites) SK- OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620 (SW480 met)	0.0	Prostate	5.3
Colon ca. HT29	0.0	Prostate ca. * (bone met)	12.8

(OD04622-01)			9060395		
Kidney Margin (OD04622-03)	4.6	3.8	Stomach Margin 9060394	0.0	0.0
Kidney Cancer (OD04450-01)	0.0	4.5	Gastric Cancer 9060397	0.0	0.0
Kidney Margin (OD04450-03)	0.0	2.3	Stomach Margin 9060396	0.0	0.0
Kidney Cancer 8120607	0.0	0.0	Gastric Cancer 064005	1.5	6.7

Table 42. Panel 4D

Tissue Name	Rel. Exp.(%) Ag548, Run 145909057	Rel. Exp.(%) Ag548, Run 145928351	Rel. Exp.(%) Ag548, Run 164886875	Tissue Name	Rel. Exp.(%) Ag548, Run 145909057	Rel. Exp.(%) Ag548, Run 145928351	Rel. Exp.(%) Ag548, Run 164886875
Secondary Th1 act	0.0	3.1	0.0	HUVEC IL-1beta	0.0	0.0	0.0
Secondary Th2 act	0.0	0.0	4.9	HUVEC IFN gamma	0.0	0.0	0.0
Secondary Tr1 act	0.0	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0	0.0
Secondary Th1 rest	0.0	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0	0.0
Secondary Th2 rest	0.0	0.0	0.0	HUVEC IL-11	0.0	0.0	0.0
Secondary Tr1 rest	0.0	7.1	0.0	Lung Microvascular EC none	0.0	2.9	0.0
Primary Th1 act	0.0	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0	0.0	0.0
Primary Th2 act	0.0	1.1	0.0	Microvascular Dermal EC none	0.0	0.0	0.0
Primary Tr1 act	0.0	0.0	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	0.0	0.0	0.0
Primary Th1 rest	0.0	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	6.5	0.0	0.0
Primary Th2 rest	0.0	0.0	0.0	Small airway epithelium none	13.4	13.6	8.8
Primary Tr1 rest	0.0	0.0	5.3	Small airway epithelium TNFalpha + IL-1beta	100.0	73.2	95.9
CD45RA CD4 lymphocyte act	0.0	0.0	0.0	Coronary artery SMC rest	0.0	0.0	0.0
CD45RO CD4 lymphocyte act	0.0	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0	0.0

[REDACTED]

PMA/ionomycin				fibroblast CCD1070 IL-1 beta			
Dendritic cells none	3.5	0.0	0.0	Dermal fibroblast IFN gamma	0.0	0.0	0.0
Dendritic cells LPS	0.0	0.0	5.4	Dermal fibroblast IL-4	0.0	0.0	0.0
Dendritic cells anti-CD40	0.0	0.0	0.0	IBD Colitis 2	3.5	2.5	19.9
Monocytes rest	0.0	0.0	0.0	IBD Crohn's	0.0	13.3	0.0
Monocytes LPS	0.0	0.0	0.0	Colon	0.0	0.0	0.0
Macrophages rest	4.5	0.0	0.0	Lung	0.0	0.0	0.0
Macrophages LPS	2.4	8.0	0.0	Thymus	0.0	0.0	0.0
HUVEC none	0.0	0.0	0.0	Kidney	87.1	100.0	100.0
HUVEC starved	0.0	0.0	0.0				

Panel 1.2 Summary: Ag548 The NOV7 gene is expressed at a very low level in this panel. The highest level of expression is seen in thymus (CT=34.33). Thus, expression of the NOV7 gene could be used to differentiate thymus tissues from other samples on this panel.

5 Furthermore, the highly specific expression pattern suggests that the NOV7 gene product may be involved in the normal homeostasis of the thymus.

Panel 1.3D Summary: Ag548 The NOV7 gene is expressed at a very low level in this panel. The highest level of expression is seen in thymus (CT=34.33). Thus, expression of the NOV7 gene could be used to differentiate thymus tissues from other samples on this panel.

10 Furthermore, the highly specific expression pattern suggests that the NOV7 gene product may be involved in the normal homeostasis of the thymus.

Panel 2D Summary: Ag548 The highest level of expression in this panel is a bladder cancer sample (CT=30.34 and 30.01) with good concordance between two independent runs.

Thus, expression of the NOV7 gene could also be used to differentiate between normal and cancerous bladder tissue. Furthermore, therapeutic modulation of the NOV7 gene product through the application of antibodies, chimeric molecules or small molecule inhibitors may be effective in the treatment of bladder cancer.

Panel 4D Summary: Ag548: This transcript is induced in small airway epithelium activated with TNFalpha+IL-1beta and is also present in normal kidney. The transcript encodes a putative serpin like molecule (serine proteinase inhibitor). Serpins participate in multiple biological processes; mutations that alter serpin function can result in pulmonary dysfunction including asthma and emphysema. Therapies designed with the protein encoded by this transcript could be important for the treatment of lung disorders such as asthma and

emphysema (Parmar and Lomas, Alpha-1-antitrypsin deficiency, the serpinopathies and conformational disease. J R Coll Physicians Lond 34(3):295-300, 2000; Malerba et al., Chromosome 14 linkage analysis and mutation study of 2 serpin genes in allergic asthmatic families. J Allergy Clin Immunol 107(4):654-814, 2001).

5 H. NOV8a – NOV8c and NOV8e: B7 family-like

Expression of the NOV8a gene (CG50309-01) and variants NOV8b (CG50309-02), NOV8c (CG50309-03), NOV8e (CG50309-05) was assessed using the primer-probe sets Ag2547 and Ag4939, described in Tables 43 and 44. Results of the RTQ-PCR runs are shown in Tables 45-50.

10 Table 43. Probe Name Ag2547

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggaggactctaattggttcatt-3'	22	1329	124
Probe	TET-5'-accttggtgctcgccctgacagt-3'-TAMRA	23	1372	125
Reverse	5'-cttcacgtcagctccagaat-3'	20	1396	126

Table 44. Probe Name Ag4939

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggaggactctaattggttcatt-3'	22	1329	127
Probe	TET-5'-accttggtgctcgccctgacagt-3'-TAMRA	23	1372	128
Reverse	5'-ttcacgtcagctccagaatc-3'	20	1395	129

Table 45. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2547, Run 206974443	Tissue Name	Rel. Exp.(%) Ag2547, Run 206974443
AD 1 Hippo	15.8	Control (Path) 3 Temporal Ctx	5.3
AD 2 Hippo	35.6	Control (Path) 4 Temporal Ctx	43.5
AD 3 Hippo	9.9	AD 1 Occipital Ctx	12.8
AD 4 Hippo	11.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	84.7	AD 3 Occipital Ctx	6.9
AD 6 Hippo	45.4	AD 4 Occipital Ctx	27.7
Control 2 Hippo	29.5	AD 5 Occipital Ctx	13.9
Control 4 Hippo	12.5	AD 5 Occipital Ctx	34.4
Control (Path) 3 Hippo	5.6	Control 1 Occipital Ctx	3.5
AD 1 Temporal Ctx	15.5	Control 2 Occipital Ctx	79.6
AD 2 Temporal Ctx	34.2	Control 3 Occipital Ctx	19.2
AD 3 Temporal Ctx	8.9	Control 4 Occipital Ctx	6.4
AD 4 Temporal Ctx	25.7	Control (Path) 1 Occipital Ctx	95.3
AD 5 Inf Temporal Ctx	74.7	Control (Path) 2 Occipital Ctx	15.9

		H596	
glioma SNB-19	0.0	Mammary gland	1.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (Fetal)	0.4	Breast ca.* (pl. ef) T47D	0.0
Heart	0.1	Breast ca. BT-549	0.0
Skeletal muscle (Fetal)	0.6	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.3
Bone marrow	0.7	Ovarian ca. OVCAR-3	0.0
Thymus	0.6	Ovarian ca. OVCAR-4	0.0
Spleen	15.4	Ovarian ca. OVCAR-5	0.0
Lymph node	0.2	Ovarian ca. OVCAR-8	0.0
Colorectal	0.2	Ovarian ca. IGROV-1	0.0
Stomach	0.1	Ovarian ca. (ascites) SK-OV-3	0.0
Small intestine	0.2	Uterus	0.4
Colon ca. SW480	0.1	Placenta	0.9
Colon ca.* SW620 (SW480 met)	0.0	Prostate	2.8
Colon ca. HT29	0.0	Prostate ca.* (bone met) PC-3	0.0
Colon ca. HCT-116	0.0	Testis	1.2
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
CC Well to Mod Diff (ODO3866)	1.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca. (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.7	Melanoma LOX IMVI	0.0
Trachea	0.7	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.1	Adipose	0.5

Table 47. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2547, Run 156849718	Tissue Name	Rel. Exp.(%) Ag2547, Run 156849718
Normal Colon	9.2	Kidney Margin 8120608	2.5
CC Well to Mod Diff (ODO3866)	2.5	Kidney Cancer 8120613	1.0
CC Margin (ODO3866)	0.6	Kidney Margin 8120614	2.1
CC Gr.2 rectosigmoid (ODO3868)	0.2	Kidney Cancer 9010320	2.9
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	5.6
CC Mod Diff (ODO3920)	0.3	Normal Uterus	2.0
CC Margin (ODO3920)	0.7	Uterine Cancer 064011	1.6
CC Gr.2 ascend colon (ODO3921)	0.9	Normal Thyroid	1.4
CC Margin (ODO3921)	1.4	Thyroid Cancer	0.8
CC from Partial Hepatectomy	1.1	Thyroid Cancer A302152	2.8

(ODO4309) Mets			
Liver Margin (ODO4309)	2.5	Thyroid Margin A302153	0.5
Colon mets to lung (OD04451-01)	0.9	Normal Breast	5.5
Lung Margin (OD04451-02)	0.7	Breast Cancer	17.0
Normal Prostate 6546-1	17.6	Breast Cancer (OD04590-01)	7.4
Prostate Cancer (OD04410)	20.4	Breast Cancer Mets (OD04590-03)	6.9
Prostate Margin (OD04410)	9.5	Breast Cancer Metastasis	100.0
Prostate Cancer (OD04720-01)	16.8	Breast Cancer	6.1
Prostate Margin (OD04720-02)	3.9	Breast Cancer	4.5
Normal Lung	4.6	Breast Cancer 9100266	9.4
Lung Met to Muscle (ODO4286)	12.9	Breast Margin 9100265	7.6
Muscle Margin (ODO4286)	3.5	Breast Cancer A209073	12.7
Lung Malignant Cancer (OD03126)	3.3	Breast Margin A2090734	3.4
Lung Margin (OD03126)	2.6	Normal Liver	0.5
Lung Cancer (OD04404)	7.1	Liver Cancer	2.7
Lung Margin (OD04404)	4.8	Liver Cancer 1025	0.5
Lung Cancer (OD04565)	1.7	Liver Cancer 1026	3.5
Lung Margin (OD04565)	1.5	Liver Cancer 6004-T	1.1
Lung Cancer (OD04237-01)	0.6	Liver Tissue 6004-N	0.6
Lung Margin (OD04237-02)	7.6	Liver Cancer 6005-T	2.5
Ocular Mel Met to Liver (ODO4310)	0.4	Liver Tissue 6005-N	0.1
Liver Margin (ODO4310)	2.2	Normal Bladder	12.1
Melanoma Metastasis	0.0	Bladder Cancer	3.0
Lung Margin (OD04321)	7.3	Bladder Cancer	2.4
Normal Kidney	5.0	Bladder Cancer (OD04718-01)	3.4
Kidney Ca, Nuclear grade 2 (OD04338)	1.7	Bladder Normal Adjacent (OD04718-03)	9.9
Kidney Margin (OD04338)	2.2	Normal Ovary	0.9
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Ovarian Cancer	3.7
Kidney Margin (OD04339)	1.5	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	11.8	Ovary Margin (OD04768-08)	0.8
Kidney Margin (OD04340)	3.7	Normal Stomach	0.5
Kidney Ca, Nuclear grade 3 (OD04348)	5.2	Gastric Cancer 9060358	0.8
Kidney Margin (OD04348)	3.7	Stomach Margin 9060359	1.3
Kidney Cancer (OD04622-01)	3.1	Gastric Cancer 9060395	3.6
Kidney Margin (OD04622-03)	0.7	Stomach Margin 9060394	1.6
Kidney Cancer (OD04450-01)	0.4	Gastric Cancer 9060397	6.2
Kidney Margin (OD04450-03)	1.6	Stomach Margin 9060396	0.3
Kidney Cancer 8120607	6.3	Gastric Cancer 064005	0.6

Table 48. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4939, Run 223789574	Tissue Name	Rel. Exp.(%) Ag4939, Run 223789574
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	12.2
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL- 1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0

Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	5.1
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	6.3
Macrophages rest	0.0	Lung	100.0
Macrophages LPS	0.0	Thymus	33.0
HUVEC none	0.0	Kidney	14.9
HUVEC starved	0.0		

Table 49. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2547, Run 156849717	Tissue Name	Rel. Exp.(%) Ag2547, Run 156849717
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1 anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0

LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	5.8
LAK cells IL-2+IL-12	0.0	Lupus kidney	6.8
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL- 1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL- 1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	67.8
Macrophages LPS	0.0	Thymus	21.6
HUVEC none	0.0	Kidney	100.0
HUVEC starved	0.0		

Table 50. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag2547, Run 171656410	Tissue Name	Rel. Exp.(%) Ag2547, Run 171656410
BA4 Control	32.5	BA17 PSP	15.8
BA4 Control2	64.6	BA17 PSP2	7.0
BA4 Alzheimer's2	8.4	Sub Nigra Control	15.6
BA4 Parkinson's	64.6	Sub Nigra Control2	13.3
BA4 Parkinson's2	100.0	Sub Nigra Alzheimer's2	6.4
BA4 Huntington's	38.4	Sub Nigra Parkinson's2	29.3
BA4 Huntington's2	16.4	Sub Nigra Huntington's	32.5
BA4 PSP	11.1	Sub Nigra Huntington's2	13.5
BA4 PSP2	31.2	Sub Nigra PSP2	3.2

BA4 Depression	16.4	Sub Nigra Depression	5.6
BA4 Depression2	7.8	Sub Nigra Depression2	2.3
BA7 Control	48.3	Glob Palladus Control	3.9
BA7 Control2	36.3	Glob Palladus Control2	3.4
BA7 Alzheimer's2	6.2	Glob Palladus Alzheimer's	4.8
BA7 Parkinson's	1.4	Glob Palladus Alzheimer's2	1.7
BA7 Parkinson's2	51.8	Glob Palladus Parkinson's	54.0
BA7 Huntington's	55.1	Glob Palladus Parkinson's2	7.1
BA7 Huntington's2	27.5	Glob Palladus PSP	2.2
BA7 PSP	29.1	Glob Palladus PSP2	2.4
BA7 PSP2	27.2	Glob Palladus Depression	2.0
BA7 Depression	7.7	Temp Pole Control	12.9
BA9 Control	30.8	Temp Pole Control2	37.9
BA9 Control2	94.0	Temp Pole Alzheimer's	8.0
BA9 Alzheimer's	2.6	Temp Pole Alzheimer's2	5.4
BA9 Alzheimer's2	20.4	Temp Pole Parkinson's	23.2
BA9 Parkinson's	38.2	Temp Pole Parkinson's2	36.6
BA9 Parkinson's2	64.2	Temp Pole Huntington's	37.6
BA9 Huntington's	45.4	Temp Pole PSP	4.0
BA9 Huntington's2	19.6	Temp Pole PSP2	6.1
BA9 PSP	14.1	Temp Pole Depression2	6.3
BA9 PSP2	6.0	Cing Gyr Control	71.7
BA9 Depression	11.0	Cing Gyr Control2	42.3
BA9 Depression2	13.2	Cing Gyr Alzheimer's	17.8
BA17 Control	43.8	Cing Gyr Alzheimer's2	4.9
BA17 Control2	57.0	Cing Gyr Parkinson's	25.5
BA17 Alzheimer's2	4.8	Cing Gyr Parkinson's2	36.9
BA17 Parkinson's	30.8	Cing Gyr Huntington's	69.7
BA17 Parkinson's2	36.1	Cing Gyr Huntington's2	16.3
BA17 Huntington's	30.1	Cing Gyr PSP	15.7
BA17 Huntington's2	12.1	Cing Gyr PSP2	2.8
BA17 Depression	11.0	Cing Gyr Depression	8.3
BA17 Depression2	20.7	Cing Gyr Depression2	9.2

CNS_neurodegeneration_v1.0 and Panel CNS_1 Summary: Ag2547 The expression of the NOV8 gene does not show any evident disease association in this panel. However, these results confirm the expression of the NOV8 gene in the brain. Please see Panel 1.3D for discussion of potential utility in the brain.

Panel 1.3D Summary: The NOV8 gene exhibits highly brain-preferential expression, with highest expression in the hippocampus (CT=26.6). The NOV8 gene encodes a protein that is homologous to a B7-family molecule. These molecules may play a role in T cell activation during the early stages of central nervous system inflammation. Inflammatory process has been proposed as an underlying etiology in Alzheimer's disease, multiple sclerosis, autism, schizophrenia and depression. Therefore, agents that inhibit the expression or action of the gene product may have utility in the treatment of these CNS disorders.

Moderate to low expression is also seen in two lung cancer cell lines (SHP-77 and NCI-H69). Thus, therapeutic modulation of the gene product could be effective in the treatment of lung cancers that were used in the derivation of these cell lines.

Among metabolic tissues, the NOV8 gene shows modest expression in the adrenal gland. Lower expression is seen in fetal skeletal muscle, fetal liver and placenta and very weak expression is seen in thyroid, pancreas, pituitary, fetal heart and adipose. Thus, modulation of the NOV8 gene or its product may be used in the therapy for metabolic disorders involving the adrenal, for example, Cushing's disease, Addison's disease etc.

The expression of the NOV8 gene is higher in skeletal muscle and liver derived from fetal sources (CTs=33-34) than in the corresponding tissues from adult sources (CTs=37-40). Thus, expression of the NOV8 gene could be used to differentiate adult and fetal liver and skeletal muscle. Furthermore, the higher levels of expression in the fetal tissues suggest that the NOV8 gene product may be involved in the development of these organs. Therefore, therapeutic modulation of the expression or function of the NOV8 gene may be effective in the treatment of diseases that affect these organs (Liu et al., Immunomodulatory effects of interferon beta-1a in multiple sclerosis. *J Neuroimmunol* 112(1-2):153-62, 2001; Omari and Dorovini-Zis, Expression and function of the costimulatory molecules B7-1 (CD80) and B7-2 (CD86) in an in vitro model of the human blood-brain barrier. *J Neuroimmunol* 113(1):129-141, 2001).

Panel 2D Summary: Ag2547 The NOV8 gene is expressed at a low level in most of the tissues in this panel. The highest level of expression is seen in a sample of breast cancer metastasis (CT=28.13). There is slightly higher expression in gastric, liver and prostate cancer than the normal adjacent tissues. Expression of the NOV8 gene could also be used to differentiate between normal and cancerous gastric, liver and prostate tissue and therapeutic modulation of the gene product could be effective in the treatment of these cancers.

Panel 4.1D Summary: Ag4939: The highest expression of the NOV8 transcript is found in lung, thymus and kidney. This transcript encodes for a B7-like molecule that is a

critical co-stimulatory molecule for T cells. Modulation of the activity of the protein encoded by this transcript by protein therapeutics or antibody could play a role in the normal homeostasis of these tissues.

- 5 **Panel 4D Summary:** Ag2547: The highest expression of the NOV8 transcript is found in kidney, lung, and thymus. This transcript encodes for a B7 like molecule that is a critical co-stimulatory molecule for T cells. Modulation of the activity of the protein encoded by this transcript by protein therapeutics or antibody could play a role in the normal homeostasis of these tissues.

I. NOV9a – NOV9d : Acyl-CoA Dehydrogenase-like

- 10 Expression of the NOV9a gene (cg-140509446/CG55900-01) and variants NOV9b (CG55900-02), NOV9c (CG55900-03) and NOV9d (CG55900-04) was assessed using the primer-probe set Ag2647, described in Table 51. Results of the RTQ-PCR runs are shown in Tables 52-55.

Table 51. Probe Name Ag2647

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aaggacttttgaaccttttct-3'	22	398	130
Probe	TET-5'-acccttagaggtgatcccgagaaaa-3'-TAMRA	26	420	131
Reverse	5'-gatgtgcatattccacatttgt-3'	22	463	132

- 15 Table 52. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2647, Run 206460425	Tissue Name	Rel. Exp.(%) Ag2647, Run 206460425
AD 1 Hippo	11.0	Control (Path) 3 Temporal Ctx	9.5
AD 2 Hippo	30.8	Control (Path) 4 Temporal Ctx	31.2
AD 3 Hippo	9.0	AD 1 Occipital Ctx	13.4
AD 4 Hippo	11.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	89.5	AD 3 Occipital Ctx	5.8
AD 6 Hippo	61.1	AD 4 Occipital Ctx	16.0
Control 2 Hippo	29.5	AD 5 Occipital Ctx	33.4
Control 4 Hippo	20.3	AD 5 Occipital Ctx	0.9
Control (Path) 3 Hippo	1.5	Control 1 Occipital Ctx	10.0
AD 1 Temporal Ctx	15.2	Control 2 Occipital Ctx	45.4
AD 2 Temporal Ctx	31.2	Control 3 Occipital Ctx	23.3
AD 3 Temporal Ctx	9.6	Control 4 Occipital Ctx	5.4
AD 4 Temporal Ctx	8.7	Control (Path) 1 Occipital Ctx	88.9
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	6.0
AD 5 Sup Temporal	62.0	Control (Path) 3	6.7

Ctx		Occipital Ctx	
AD 6 Inf Temporal Ctx	35.8	Control (Path) 4 Occipital Ctx	23.2
AD 6 Sup Temporal Ctx	51.1	Control 1 Parietal Ctx	19.1
Control 1 Temporal Ctx	8.1	Control 2 Parietal Ctx	44.4
Control 2 Temporal Ctx	39.2	Control 3 Parietal Ctx	9.2
Control 3 Temporal Ctx	1.8	Control (Path) 1 Parietal Ctx	52.1
Control 3 Temporal Ctx	7.5	Control (Path) 2 Parietal Ctx	26.2
Control (Path) 1 Temporal Ctx	53.6	Control (Path) 3 Parietal Ctx	7.5
Control (Path) 2 Temporal Ctx	40.3	Control (Path) 4 Parietal Ctx	40.6

Table 53. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2647, Run 164023714	Tissue Name	Rel. Exp.(%) Ag2647, Run 164023714
Liver adenocarcinoma	14.9	Kidney (fetal)	8.4
Pancreas	3.1	Renal ca. 786-0	5.8
Pancreatic ca. CAPAN 2	2.9	Renal ca. A498	7.1
Adrenal gland	6.3	Renal ca. RXF 393	3.9
Thyroid	12.4	Renal ca. ACHN	13.1
Salivary gland	4.4	Renal ca. UO-31	6.3
Pituitary gland	6.8	Renal ca. TK-10	4.3
Brain (fetal)	4.5	Liver	4.9
Brain (whole)	7.3	Liver (fetal)	2.1
Brain (amygdala)	6.5	Liver ca. (hepatoblast) HepG2	4.4
Brain (cerebellum)	5.3	Lung	2.9
Brain (hippocampus)	7.4	Lung (fetal)	9.9
Brain (substantia nigra)	1.2	Lung ca. (small cell) LX-1	4.3
Brain (thalamus)	3.8	Lung ca. (small cell) NCI-H69	2.7
Cerebral Cortex	23.8	Lung ca. (s.cell var.) SHP-77	15.1
Spinal cord	14.3	Lung ca. (large cell) NCI-H460	3.7
glio/astro U87-MG	19.3	Lung ca. (non-sm. cell) A549	5.4
glio/astro U-118-MG	6.8	Lung ca. (non-s.cell) NCI-H23	15.2
astrocytoma SW1783	14.1	Lung ca. (non-s.cell) HOP-62	10.1
neuro*; met SK-N-AS	6.2	Lung ca. (non-s.cl) NCI-H522	4.3
astrocytoma SF-539	7.5	Lung ca. (squam.) SW 900	4.0
astrocytoma SNB-75	2.4	Lung ca. (squam.) NCI-H596	3.0

100-443886-1000

100-443886-1000

Tissue Name	Rel. Exp.(%) Ag2647, Run 162554730	Tissue Name	Rel. Exp.(%) Ag2647, Run 162554730
Secondary Th1 act	8.7	HUVEC IL-1beta	1.7
Secondary Th2 act	20.3	HUVEC IFN gamma	7.6
Secondary Tr1 act	10.7	HUVEC TNF alpha + IFN gamma	5.2
Secondary Th1 rest	6.4	HUVEC TNF alpha + IL4	3.5
Secondary Th2 rest	6.5	HUVEC IL-11	4.1
Secondary Tr1 rest	12.2	Lung Microvascular EC none	7.8
Primary Th1 act	15.4	Lung Microvascular EC TNFalpha + IL-1beta	7.1
Primary Th2 act	15.2	Microvascular Dermal EC none	11.6
Primary Tr1 act	23.2	Microvascular Dermal EC TNFalpha + IL-1beta	6.6
Primary Th1 rest	34.9	Bronchial epithelium TNFalpha + IL1beta	2.6
Primary Th2 rest	21.5	Small airway epithelium none	1.9
Primary Tr1 rest	27.9	Small airway epithelium TNFalpha + IL-1beta	15.7
CD45RA CD4 lymphocyte act	7.6	Coronary artery SMC rest	14.1
CD45RO CD4 lymphocyte act	12.3	Coronary artery SMC TNFalpha + IL-1beta	5.2
CD8 lymphocyte act	9.0	Astrocytes rest	11.8
Secondary CD8 lymphocyte rest	16.0	Astrocytes TNFalpha + IL-1beta	4.3
Secondary CD8 lymphocyte act	6.0	KU-812 (Basophil) rest	10.4
CD4 lymphocyte none	11.7	KU-812 (Basophil) PMA/ionomycin	20.7
2ry Th1/Th2/Tr1 anti- CD95 CH11	12.7	CCD1106 (Keratinocytes) none	7.6
LAK cells rest	22.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	15.8	Liver cirrhosis	3.7
LAK cells IL-2+IL-12	11.9	Lupus kidney	3.2
LAK cells IL-2+IFN gamma	23.7	NCI-H292 none	40.1
LAK cells IL-2+ IL-18	23.7	NCI-H292 IL-4	28.9
LAK cells PMA/ionomycin	2.6	NCI-H292 IL-9	28.9
NK Cells IL-2 rest	7.9	NCI-H292 IL-13	15.0
Two Way MLR 3 day	21.5	NCI-H292 IFN gamma	18.3
Two Way MLR 5 day	5.4	HPAEC none	5.0
Two Way MLR 7 day	3.8	HPAEC TNF alpha + IL-1 beta	4.0
PBMC rest	2.9	Lung fibroblast none	13.8
PBMC PWM	22.4	Lung fibroblast TNF alpha + IL- 1 beta	8.6
PBMC PHA-L	12.1	Lung fibroblast IL-4	15.5
Ramos (B cell) none	8.5	Lung fibroblast IL-9	15.8
Ramos (B cell) ionomycin	28.1	Lung fibroblast IL-13	8.5

B lymphocytes PWM	34.9	Lung fibroblast IFN gamma	18.4
B lymphocytes CD40L and IL-4	21.3	Dermal fibroblast CCD1070 rest	14.7
EOL-1' dbcAMP	10.8	Dermal fibroblast CCD1070 TNF alpha	31.4
EOL-1 dbcAMP PMA/ionomycin	10.6	Dermal fibroblast CCD1070 IL-1 beta	7.7
Dendritic cells none	12.4	Dermal fibroblast IFN gamma	6.8
Dendritic cells LPS	7.7	Dermal fibroblast IL-4	8.3
Dendritic cells anti-CD40	12.2	IBD Colitis 2	0.2
Monocytes rest	23.7	IBD Crohn's	1.9
Monocytes LPS	5.6	Colon	33.4
Macrophages rest	19.5	Lung	12.1
Macrophages LPS	4.4	Thymus	100.0
HUVEC none	8.0	Kidney	29.3
HUVEC starved	15.6		

CNS_neurodegeneration_v1.0 Summary: Ag2647 While the expression of the NOV9 gene does not show any evident disease association, these results confirm the expression of the NOV9 gene in the brain.

5 **Panel 1.3D Summary:** Ag2647 The NOV9 gene is expressed ubiquitously among the samples in this panel, with highest levels in fetal skeletal muscle (CT=28.7). Moderate levels or expression are detected in the kidney and ovary. Lower but still significant levels are seen in a variety of metabolic tissues, including skeletal muscle, heart, thyroid, adrenal, placenta, pituitary, pancreas, fetal liver and adipose. The NOV9 gene encodes a protein with homology
10 to Acyl-CoA dehydrogenase and may be involved in fatty acid metabolism. In addition, small molecule therapeutics directed against the NOV9 gene product may be useful in metabolic disorders such as obesity.

Higher levels of expression in fetal skeletal muscle (CT=28.7) when compared to adult skeletal muscle (CT=31.9) suggest that expression of the NOV9 gene could be used to
15 differentiate between the two sources of tissue. Furthermore, the higher levels of expression in fetal skeletal muscle suggest that the protein encoded by the NOV9 gene may enhance muscular growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by the NOV9 gene could be useful in treatment of muscle related diseases. More specifically, treatment of weak
20 or dystrophic muscle with the protein encoded by the NOV9 gene could restore muscle mass or function.

Panel 2D Summary: Ag2647 The NOV9 gene is expressed at a moderate to low levels in most of the samples in this panel, with highest expression in normal kidney

(CT=28.43). There is slightly increased expression in gastric, ovarian, breast, lung and colon cancers when compared to normal adjacent tissues from these organs. Thus, expression of the NOV9 gene could potentially be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition of the activity of the protein encoded by the NOV9 gene through the application of small molecule inhibitors may be beneficial for the treatment of these cancers.

Panel 4D Summary: Ag 2647 This transcript is expressed at high levels in the thymus (CT 28.4), a muco-epidermoid cell line (H292), activated normal B cells (B cells stimulated with PWM) , a lymphoma B cell line (Ramos) and primary Th1 cells. The NOV9 gene is expressed at lower but still significant levels in a wide range of cell types with significance in the immune response in health and disease. Therefore, inhibition of the function of the protein encoded by the NOV9 gene through the application of a small molecule drug may block the functions of B cells, T cells, and other cell types such as mucus producing cells, lung fibroblasts. This inhibition could potentially lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases, including asthma, chronic obstructive pulmonary disease, allergies, inflammatory bowel disease, lupus erythematosus, and rheumatoid arthritis.

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims.